

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 1999, 16:56:23 ; Search time 15.18 Seconds
(without alignments)
3845.548 Million cell updates/sec

Title: US-09-001-039a-47

Sequence: 1 MQIEISTCFFLCLRLRCFSA.....WVHQIALRMEVLGCEADLY 1457

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database :

PIR_60:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7340	94.1	2351	1 E2HU	coagulation factor
2	6261	80.3	2319	2 A47004	coagulation factor
3	2391.5	30.7	2224	1 KEHUS	coagulation factor
4	2370	30.4	2211	1 KEBO5	coagulation factor
5	1745	22.4	1069	1 KUHU	ferroxidase (EC 1.1.1.1)
6	1697	21.8	1059	2 A35210	ferroxidase (EC 1.1.1.1)
7	1104	14.2	216	2 A44258	factor VIII-associated
8	663	8.5	427	2 JC4915	aggs protein precursor
9	657	8.4	463	2 A36479	milk fat globule m
10	635	8.1	401	2 S65138	glycoprotein antiq
11	635	8.1	427	2 S74211	PAS-6/7 protein pr
12	543	7.0	869	2 A25945	coagulation factor
13	443	5.7	927	2 JQ0948	A5 antigen precursor
14	424.5	5.4	218	2 A47285	milk fat globule p
15	306.5	3.9	313	2 S52093	hemocytin - silkwo
16	262	3.4	845	3 JC5256	adipocyte transcri
17	221	2.8	719	2 S51739	transcription repr
18	193	2.5	913	2 A48280	receptor tyrosine
19	192	2.5	876	2 A49508	protein-tyrosine k
20	185.5	2.4	855	2 S42621	protein-tyrosine k
21	185.5	2.4	910	2 A53137	tyrosine kinase re
22	173.5	2.2	819	2 I48859	tyro 10 receptor k
23	149.5	1.9	578	2 S66353	L-ascorbate oxidas
24	148	1.9	622	2 S62580	hypothetical prote
25	141.5	1.8	1196	2 S46430	botulinum neurotox
26	140.5	1.8	1196	2 JQ1467	toxin, nontoxic co
27	139	1.8	1193	2 JC4901	nontoxic nonhemag
28	133.5	1.7	2925	3 T00133	RNA-directed RNA p
29	132.5	1.7	1159	2 S22768	130K protein - mal
30	130	1.7	818	1 A30107	dipeptidyl aminope
31	128.5	1.6	994	2 S00960	hypothetical prote
32	127	1.6	1274	2 T02636	DL protein homolog
33	127	1.6	1333	2 A37488	Ras guanine nucleo
34	125	1.6	2167	2 S50658	bud emergence prot
35	125	1.6	2368	2 S46005	ESR1 protein - yea
36	124	1.6	554	2 T01240	probable diphenol
37	124	1.6	3169	2 T00296	toxin B - Escheric
38	123	1.6	587	1 KSKVAO	L-ascorbate oxidas
39	123	1.6	1471	1 S30790	myosin MYO4 - yeas

BEST AVAILABLE COPY

40	122.5	1.6	1178	2 S54073	probable membrane
41	121.5	1.6	624	2 A36962	laccase (EC 1.10.3
42	121.5	1.6	2216	2 S78398	hypothetical prote
43	121	1.6	767	2 F64605	iron(III) dicitrat
44	120.5	1.5	1405	1 DJ2PA	DNA-directed DNA p
45	120.5	1.5	2139	2 S46404	vitellinogenin - yel

ALIGNMENTS

RESULT 1
E2HU
coagulation factor VIII precursor - human
N:Alternate names: antihemophilic factor A; coagulation factor VIIIC; procoagulant co
C:Species: Homo sapiens (man)
C>Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 26-Feb-1999
C:Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S66445;
R:Gitschier, J.; Wood, W.I.
Hum. Mol. Genet. 1, 199-200, 1992
A:Title: Sequence of the exon-containing regions of the human factor VIII gene.
A:Reference number: I54318; MUID:93265012
A:Accession: I54318
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1921, 'S', 1923-2351 <RES>
A:Cross-references: GB:M88648; NID:g182381; PID:g182383
R:Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschier, J.; Keyt, B.; Seeb
Nature 312, 330-337, 1984
A:Title: Expression of active human factor VIII from recombinant DNA clones.
A:Reference number: A00525; MUID:85061548
A:Accession: A00525
A:Molecule type: mRNA
A:Residues: 1-2351 <MOO>
A:Cross-references: EMBL:X01165; EMBL:X01166; EMBL:X01179
R:Toole, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.L.; Pittman, D.D
s, D.N.; Hewick, R.M.
Nature 312, 342-347, 1984
A:Title: Molecular cloning of a cDNA encoding human antihemophilic factor.
A:Reference number: I58059; MUID:85061550
A:Accession: I58059
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-74, 'V', 76-1259, 'E', 1261-2351 <RE2>
A:Cross-references: GB:K01740; NID:g182802; PID:g182803
R:Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.; K
.B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen, J.;
DNA 4, 333-349, 1985
A:Title: Characterization of the polypeptide composition of human factor VIII:C and t
A:Reference number: A23584; MUID:86081164
A:Accession: A23584
A:Molecule type: mRNA
A:Residues: 1-2351 <TRU>
A:Cross-references: GB:M14113; NID:g182817; PID:g182818
R:Eaton, D.; Rodriguez, H.; Vohar, G.A.
Biochemistry 25, 505-512, 1986
A:Title: Proteolytic processing of human factor VIII. Correlation of specific cleavag
ity.
A:Reference number: A26174; MUID:86159740
A:Accession: A26174
A:Molecule type: protein
A:Residues: 20-36; 392-399, 'X', 401-402; 1668-1678; 1709-1722, 'D', 1723-1725; 1741-1755 <EA
R:Pittman, D.D.; Wang, J.H.; Kaufman, R.J.
Biochemistry 31, 3315-3325, 1992
A:Title: Identification and functional importance of tyrosine sulfate residues within
A:Reference number: A42348; MUID:92207952
A:Accession: A42348
A:Molecule type: protein
A:Residues: 20-36; 356-371; 392-408; 582-594; 1668-1669, 'X', 1671; 1672-1692; 1693-1708; 1709
A:Experimental source: recombinant material from Chinese hamster ovary cells
A>Note: sequence extracted from NCBI backbone and corrected to correspond with the pu
R:Ray, P.J.; Smudzyn, T.M.
J. Biol. Chem. 264, 14005-14010, 1989

A>Title: Intersubunit fluorescence energy transfer in human factor VIII.
A:Reference number: A43986; MUID:89340500
A:Accession: A43986
A:Molecule type: protein
A:Residues: 'X',517-523;1853-1860,'X',1862-1864,'X',1866 <FAY>
R:leyte, A.; van Schijndel, H.B.; Niehrs, C.; Huttner, W.B.; Verbeet, M.P.; Mertens, K.;
J. Biol. Chem. 266, 740-746, 1991
A>Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for
A:Reference number: A56109
A:Contents: annotation; sulfation
R:Gitschier, J.; Wood, W.I.; Goralka, T.M.; Wion, K.L.; Chen, E.Y.; Eaton, D.H.; Vehar,
Nature 312, 326-330, 1984
A>Title: Characterization of the human factor VIII gene.
A:Reference number: A56196; MUID:85061547
A:Contents: annotation; introns
R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.
Protein Sci. 4, 740-746, 1995
A>Title: Locations of disulfide bonds and free cysteines in the heavy and light chains of
A:Reference number: A56216; MUID:95338127
A:Contents: annotation; disulfide bonds
R:Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulfhydryls
R:Kjalke, M.; Heding, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M.
Eur. J. Biochem. 234, 773-779, 1995
A>Title: Amino acid residues 721-729 are required for full factor VIII activity.
A:Reference number: S63527; MUID:96163459
A:Accession: S63527
A:Molecule type: protein
A:Residues: 733-752;753-759 <KJA>
R:Lind, P.; Larsson, K.; Sjira, J.; Sydow-Baekman, M.; Almstedt, A.; Gray, E.; Sandberg
Eur. J. Biochem. 232, 19-27, 1995
A>Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction
A:Reference number: S66445; MUID:96048024
A:Accession: S66445
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1668-1685 <LIN>
C:Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure pro-
C:Genetics:
A:Gene: GDB:F8C
A:Cross-references: GDB:119124; OMIM:306700
A:Map position: Xq28-Xq28
A:Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro-
A:Pathway: blood coagulation
A:Family: blood coagulation factor VIII; discoïdin I amino-terminal homology; ferroxidase
A:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla-
A:Domain: signal sequence #status predicted <SIG>
F:20-2351/Product: coagulation factor VIII #status experimental <MAT>
F:20-740/Product: coagulation factor VIIa heavy chain #status experimental <ACH>
F:20-356/Domain: A1 <DA1>
F:23-348/Domain: ferroxidase repeat homology <FO1>
F:392-759/Domain: A2 <DA2>
F:402-730/Domain: ferroxidase repeat homology <FO2>
F:760-1667/Domain: B <DB0>
F:1668-2351/Product: coagulation factor VIIa light chain #status experimental <ACL>
F:1709-2038/Domain: A3 <DA3>
F:1716-2038/Domain: ferroxidase repeat homology <FO3>
F:2039-2191/Domain: C1 <DC1>
F:2039-2191/Domain: discoïdin I amino-terminal homology <DN1>
F:2039-2188/Domain: discoïdin I amino-terminal homology <DN1>
F:2192-2351/Domain: C2 <DC2>
F:2192-2345/Domain: discoïdin I amino-terminal homology <DN2>
F:60,258,601,776,803,847,919,962,982,1020,1024,1074,1085,1204,1274,1278,1301,1319,1403,1
F:172-198,267-348,547-573,649-730,1851-1877,1918-1922,2040-2188/Disulfide bonds: #status
F:355-356/Cleavage site: Arg-Met (coagulation factor Xa, protein C) #status predicted
F:365,737,738,742,1683,1699/Binding site: sulfate (Tyr) (covalent) #status experimental
F:391-392/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:414,426/Binding site: sulfate (Tyr) (covalent) #status predicted
F:759-760/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:1667-1668/Cleavage site: Arg-Glu (unidentified proteinase) #status experimental
F:1708-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:1740-1741/Cleavage site: Arg-Ala (coagulation factor Xa) #status experimental
F:2193-2345/Disulfide bonds: #status predicted

Query Match	94.18;	Score 7340;	DB 1;	Length 2351;
Best Local Similarity	62.0%;	Pred. No. 0;		
Matches 1457;	Conservative 0;	Mismatches 0;	Indels 894;	Gaps 1;
QY 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSWDMQSDIGELLPVDAFPFPPVPSFPFN 60				
Db 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSWDMQSDIGELLPVDAFPFPPVPSFPFN 60				
QY 61 TSVYKKTLVEFTDHLFNIAKPRPPMGLGPTIQAEVYDTVITLKNASHPVSLHAV 120				
Db 61 TSVYKKTLVEFTDHLFNIAKPRPPMGLGPTIQAEVYDTVITLKNASHPVSLHAV 120				
QY 121 GVSYWKASEGAEYDDQTSQREKEDDKVFPGSGHTYVQVLKENGPMASDPLCLTYSLSH 180				
Db 121 GVSYWKASEGAEYDDQTSQREKEDDKVFPGSGHTYVQVLKENGPMASDPLCLTYSLSH 180				
QY 181 VDLVKDLNSGLIGALLVCRESGLAKEKTQTILAKFILLFAVDEGKSWHSETKNSLMQDRD 240				
Db 181 VDLVKDLNSGLIGALLVCRESGLAKEKTQTILAKFILLFAVDEGKSWHSETKNSLMQDRD 240				
QY 241 AASARAWPKMHTVNGVYVNRSLPGLIGCHRSYVHWYVIGMGTTPVHSIFLEGHTFLVRNH 300				
Db 241 AASARAWPKMHTVNGVYVNRSLPGLIGCHRSYVHWYVIGMGTTPVHSIFLEGHTFLVRNH 300				
QY 301 ROASLEISPTITLTAQTLMDIGQFLFCHTISHQHDGMEAYVKVDSCEEPQLRMKNNE 360				
Db 301 ROASLEISPTITLTAQTLMDIGQFLFCHTISHQHDGMEAYVKVDSCEEPQLRMKNNE 360				
QY 361 EAEYDDDLTDEMDVVRFDNNSPSFIQIRSYAKKHPTWVHYIAAEEDMDYAPLVLA 420				
Db 361 EAEYDDDLTDEMDVVRFDNNSPSFIQIRSYAKKHPTWVHYIAAEEDMDYAPLVLA 420				
QY 421 PDDRSYKSQYLNNGPQIRGRKKYKVRFMAYTDETFKTRALQHESGILGPLYGEVGDTL 480				
Db 421 PDDRSYKSQYLNNGPQIRGRKKYKVRFMAYTDETFKTRALQHESGILGPLYGEVGDTL 480				
QY 481 LIIFKNQASRPYNTYPHGITDVVRPLYSRRLPKGVKHLKDFILPGEIFKYYKWTVVEDGP 540				
Db 481 LIIFKNQASRPYNTYPHGITDVVRPLYSRRLPKGVKHLKDFILPGEIFKYYKWTVVEDGP 540				
QY 541 TKSDPRLCTRYYSFVNMERDLASGLIGPLLCYKESVDQRGNOIMSDKRNVILFSVFDE 600				
Db 541 TKSDPRLCTRYYSFVNMERDLASGLIGPLLCYKESVDQRGNOIMSDKRNVILFSVFDE 600				
QY 601 NRSWYLTENIQRFILPNPAGVQLDEBEFOASNIMHSINGVYVFDLSQVCLHEVAYWYILS 660				
Db 601 NRSWYLTENIQRFILPNPAGVQLDEBEFOASNIMHSINGVYVFDLSQVCLHEVAYWYILS 660				
QY 661 IGAQTDFLSVFFSGYTEFKHKMYEDTLTLFPFSGEYVFMSEMPGLIIGCHNSDERNRG 720				
Db 661 IGAQTDFLSVFFSGYTEFKHKMYEDTLTLFPFSGEYVFMSEMPGLIIGCHNSDERNRG 720				
QY 721 MTALLKVSCKDKNTGDIYEDSYEDISAYLLSKNNAIEPRSF----- 762				
Db 721 MTALLKVSCKDKNTGDIYEDSYEDISAYLLSKNNAIEPRSFSONSRHPSTRQKQFNATTI 780				
QY 762 ----- 762				
Db 781 PENDIEKTDPMFAHRTMPMKIQNVSSDDLMLLROSPTPHGLSLSDLQEAKEYEPPSDPS 840				
QY 762 ----- 762				
Db 841 PGAIDSNNSLSEMTHERPOLHSGDMVFTPESGIQLRLNEKLGTTAATELKKLDFRVSST 900				
QY 762 ----- 762				
Db 901 SNNLISITPSDNLAAGTDNTSSIGPPSPMVHYDSQLDTLFGKKSPLTESGPISTSEE 960				
QY 762 ----- 762				
Db 961 NNDSKLLSEGLMNSQESSWGKNVSSSTESGRLFKGKRAHGPALLTKDNALLFKVISILKTN 1020				

QY 762 ----- 762
Db 1021 KTSNNSATNRKTHIDBPSLLIENSPVQNILSDTEFKKVTPLIHDRMLMDKNATALRL 1080
QY 762 ----- 762
Db 1081 NHMSNKTSSKNMEMVQOKKEGPIDPAQNPDMSEFKMLFLPESARNIQRTHGKNSLSG 1140
QY 762 ----- 762
Db 1141 QGSPKQVLVSLGPEKSVEGQNFLSEKNKVVVGKEFTKDVGLKEMVFPSSRNLFLTNDN 1200
QY 762 ----- 762
Db 1201 LHENHNQEKKIQEIEKETLIQENVVLPQIHVYGTKNFMKNLEFLSTRQNVESYD 1260
QY 762 ----- 762
Db 1261 GAYAPVLDQFRSLNDSTNRTKHTAHPSKKEEENLEGLGNQTKQIYEKYACTTRISPNT 1320
QY 762 ----- 762
Db 1321 SQQNFVQRSKRALKQFRLPLEETELEKRIIVDDTSTOWSKNMHLTPSTLTQIDYNEKE 1380
QY 762 ----- 762
Db 1381 KGATQSPSLDCLTRSHSIPQANRSPPIAKVSSPSPRIYLTRVLEQDNSSHLPASYS 1440
QY 762 ----- 762
Db 1441 RKKDSGVQESSHFLQAKKNNLSLILTEMTGDQREVSLGTSATNSVTYKKVENTVLP 1500
QY 762 ----- 762
Db 1501 KPDLPTSGKVELLPKVHIYQKDLFPETNSGSPGHLDLVEGSLQTEGAIKWNEANRP 1560
QY 762 ----- 762
Db 1561 GKVPFLRVATESAKTPSKLLDPLAMDNHGTOIPKEEWSQEKSPERTAFAKKKDTIISL 1620
QY 762 -----SONPPVLKRHQREITRTTLOSQEE 786
Db 1621 NACESNHAIAINEGQNKPEIEVTWAKQGRTERLCSQNPVLKRHQREITRTTLOSQEE 1680
QY 787 IDYDITISVEKKEDFDIYDEDENSPSPFOKTRHYFIAAVERLWDYGMSSPHVLNR 846
Db 1681 IDYDITISVEKKEDFDIYDEDENSPSPFOKTRHYFIAAVERLWDYGMSSPHVLNR 1740
QY 847 AQSQSVQFQKVVFOETDGSFTQPLRYGELNEHLGLGPIYIRAEVEDNIMVTFRNQASR 906
Db 1741 AQSQSVQFQKVVFOETDGSFTQPLRYGELNEHLGLGPIYIRAEVEDNIMVTFRNQASR 1800
QY 907 PYSFYSSLISYEEDQROGAEPKRNFKVKNETKTYFWKVQHMAPTKDEPDKAWAYFSDV 966
Db 1801 PYSFYSSLISYEEDQROGAEPKRNFKVKNETKTYFWKVQHMAPTKDEPDKAWAYFSDV 1860
QY 967 DLEKDVHSGLIGPLVCHTNTLNPAHQVTVQEFALFFTIFDETSMVTFTEMERNCRA 1026
Db 1861 DLEKDVHSGLIGPLVCHTNTLNPAHQVTVQEFALFFTIFDETSMVTFTEMERNCRA 1920
QY 1027 PCNIQMEDPTEKENYRFAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSHFSGH 1086
Db 1921 PCNIQMEDPTEKENYRFAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSHFSGH 1980
QY 1087 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVVECLIGEHLHAGSTLFLVYSNC 1146
Db 1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVVECLIGEHLHAGSTLFLVYSNC 2040
QY 1147 QTPLGMSGHIRDFQITASGQYGOWAPKLARLHYSGSINAMSTKEPFSWIKVDLAPMI 1206
Db 2041 QTPLGMSGHIRDFQITASGQYGOWAPKLARLHYSGSINAMSTKEPFSWIKVDLAPMI 2100

QY 1207 HGKTQGARQKFSLLYSQFILMYSLDGKKMQTYRGNSTGTLMAVFEQNVDSGCIKHNIEN 1266
Db 2101 HGKTQGARQKFSLLYSQFILMYSLDGKKMQTYRGNSTGTLMAVFEQNVDSGCIKHNIEN 2160
QY 1267 PPIIARYIRLHPHTHSIRSTLRMEIMGCDLNSCMLPGMESKAISDAQITASSYFTNMA 1326
Db 2161 PPIIARYIRLHPHTHSIRSTLRMEIMGCDLNSCMLPGMESKAISDAQITASSYFTNMA 2220
QY 1327 TWSPSKARLHLOGSNAMRPQVNNPKEMLOQVDFOKTMKVTGVTQGVKSLLTSMYKEFL 1386
Db 2221 TWSPSKARLHLOGSNAMRPQVNNPKEMLOQVDFOKTMKVTGVTQGVKSLLTSMYKEFL 2280
QY 1387 ISSQDGHQWTLFQNGKVKVFOGNQDSFTPVVNSLDPPLLTRYLRIHPQSWHQTALRM 1446
Db 2281 ISSQDGHQWTLFQNGKVKVFOGNQDSFTPVVNSLDPPLLTRYLRIHPQSWHQTALRM 2340
QY 1447 EVLGCEAQDLY 1457
Db 2341 EVLGCEAQDLY 2351

RESULT 2
A47004
coagulation factor VIII precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 31-Oct-1997
C:Accession: A47004
R:Elder, B.; Iakich, D.; Gitschler, J.
Genomics 16, 374-379, 1993
A:Title: Sequence of the murine factor VIII cDNA.
A:Reference number: A47004; MUID:93300511
A:Accession: A47004
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2319 <ELD>
A:Cross-references: GB:L05573; NID:g192456; PID:g192457
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxi
F;1-19/Domain: signal sequence #status predicted <SIG>
F;23-349/Domain: ferroxidase repeat homology <FO1>
F;1686-2006/Domain: ferroxidase repeat homology <FO2>
F;1686-2006/Domain: ferroxidase repeat homology <FO3>
F;2007-2156/Domain: discoidin I amino-terminal homology <DN1>
F;2160-2313/Domain: discoidin I amino-terminal homology <DN2>

Query Match 80.3%; Score 6261; DB 2; Length 2319;
Best Local Similarity 53.5%; Pred. No. 0;
Matches 1244; Conservative 97; Mismatches 109; Indels 876; Gaps 9;
QY 1 MQEISTCEFLCLRCFSATRRXYLGAVELSWDMQSD-IGELPYDARFPPRPVRSFPE 59
Db 1 MQIALFACFFLSLFNCSSAIRRYLGAVELSWNYIQSDLLSVLHDSRFLPRMSTSFPE 60
QY 60 NTSVYVKKTLFVEFTDHLFNIAKPRPMMGLGPTIQAEVYDVTYVITLKNMASHPVSLHA 119
Db 61 NTSIMYKKTVFVEYKQQLFNIAKPRPMMGLGPTIMTEVHDTVITLKNMASHPVSLHA 120
QY 120 VGVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTFYSYL 179
Db 121 VGVSYWKASEGDEYEDQTSQMEKEDDKVFPGESHYVWQVLKENGPMASDPLCLTFYSYL 180
QY 180 HVDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFTLLFAVFDEGKSWHSETKNSLMQDR 239
Db 181 HVDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFTLLFAVFDEGKSWHSETKNSLMQDR 240
QY 240 DAASARAWPKMHTVNGVYVNRSLPGLIGCHRSVYMHVIGMGTTPREYHSIFLEGHTFLVRN 299
Db 241 DSASARDWPKMHTVNGVYVNRSLPGLIGCHRSVYMHVIGMGTTPREYHSIFLEGHTFLVRN 300
QY 300 HROASLEISPTITFLAQTLLMDLGFLLFCHISSHQHDGMEAYVYVDSCPREPQLRMK-N 358
Db 301 HROASLEISPTITFLAQTLLMDLGFLLFCHISSHQHDGMEAYVYVDSCPREPQLRMKNN 360

QY 359 NEAEDYDDDLTDESEMDVVRFDNDDSSPFIQIRSYAKKHPTWVHYIAAEEEDWDYAPLV 418
 Db 361 NEEMEDYDDDL - SEMDMFTLDYDSSP - FIOIRSYAKKYPKTWVHYISAEEDWDYAPSV 418
 QY 419 LAPDRSYKSOYLNGPQRIGRKYKVRFMAYTDEFTKREAOIHESGLIPPLYGEVGD 478
 Db 419 PTDNGSYKSOYLSNGPHRIGRKYKVRFIAYTDEFTKRETIQIHESGLIPPLYGEVGD 478
 QY 479 TLLIFKNQASRPYNTYPHGITDVRPLYSRRLPKGVKHLKDFPLIGEIEFKYKWTVED 538
 Db 479 TLLIFKNQASRPYNTYPHGITDVRPLYSRRLPKGVKHLKDFPLIGEIEFKYKWTVED 538
 QY 539 GPTKSDPRCLTRYSSFEVNMERDLASGLIGPLLCYKESVDQRGNOIMSDKRNVLFSVF 598
 Db 539 GPTKSDPRCLTRYSSFEVNMERDLASGLIGPLLCYKESVDQRGNOIMSDKRNVLFSVF 598
 QY 599 DENRSWYLTENIORFLPNPAGVQLDEDEFOASNIMHSINGYVFDISLOLVCLEHVAWYI 658
 Db 599 DENRSWYLTENIORFLPNPAGVQLDEDEFOASNIMHSINGYVFDISLOLVCLEHVAWYI 658
 QY 659 LSIGAQTDFLSVFFSGYTFKHKMVEEDTLTLPFSGETVEMSMENPGMLILGCHNSDFRN 718
 Db 659 LSIGAQTDFLSVFFSGYTFKHKMVEEDTLTLPFSGETVEMSMENPGMLILGCHNSDFRN 718
 QY 719 RGMFTALLKYSSCDKNTGDIYEDYSEDIASAYLLSKNNAIEPRFSQN----- 765
 Db 719 RGMFTALLKYSSCDKNTGDIYEDYSEDIASAYLLSKNNAIEPRFSQN----- 765
 QY 765 ----- 765
 Db 779 TIPKNDMEKIEPOFEEIAEMLKVQSVSVSDMLMLLGQSHPTPHGLFLSDGQEAIIYEAJHD 838
 QY 765 ----- 765
 Db 839 DHSPNAIDSNEGPSKVTOQLRPESHSEKIVFTPOPGLQIRSNKSLETTIEVKWKGLGLOY 898
 QY 765 ----- 765
 Db 899 SLSPLNMTTILSDNLKATFEKTDSSGFPDMPVHSSSKLSTAFGKKAYSLVGSHPVPLN 958
 QY 765 ----- 765
 Db 959 ASENSDSNILDSTIMYSQESLPRDNILSIENDRLREKRFHGIALTKONTLFFKDNVSL 1018
 QY 765 ----- 765
 Db 1019 MKTNKTYNHSITNEKLHTESPTSIENSTTDIQAALIKVNSEIQEVTALIHGTLGKNST 1078
 QY 765 ----- 765
 Db 1079 YLRNLHMLNRTSTKKNDFHRRKDEDPIDODENTIMPFSKMLFLSESSNWEKKTNGNNS 1138
 QY 765 ----- 765
 Db 1139 LNSEQEHPKQVLYLMFKKYKNQSFLEKKNVYEQDFTKNIGLKDMAFPNMSIFLT 1198
 QY 765 ----- 765
 Db 1199 TISNVHENGHNQEKNIQEEIEKALIEEKVLPQVHEATGSKNFLKDILILGTRQNISL 1258
 QY 765 ----- 765
 Db 1259 YEYHVPVLQNTITSINSTNTVQIHNEHFVKRRKDETNSEGLVNKTREMVKNYPSQKNIT 1318
 QY 765 ----- 765
 Db 1319 TQSKRALQGFRLSTOWLKTINCSIOCIKQIDHSKEMKFFITKSSLSLSSSVIKSTQTN 1378
 QY 765 ----- 765
 Db 1379 SSSDHIWKTSAFPPIDLKRSPFONKFSHVQASSYIYDFKTKSSRIOESNNFLKETKINNP 1438
 QY 765 ----- 765

Db 1439 SLAIPMNWFIQGFRTSPGKSNNTNSYTKKRENIIFLKPRTLPEESGKIELLPQVSIQEE 1498
 QY 765 ----- 765
 Db 1499 EILPTEHSGSPGHLNLMKEVFLQKIQGPTKWNKAKRHGESIKKTESSKNTRSKLLNH 1558
 QY 765 ----- 765
 Db 1559 AMDYHYAAQIPKDMWKSKEKSPEIISIKQEDTILSLRPHGNSHSIGANEKQNPQRETW 1618
 QY 765 ----- 765
 Db 1619 VKQOTORTCSQIPVLRKHQREL - SAFQSEQEAATDYDAITIE - TIEFDIYSEDIKO 1675
 QY 812 SPRSFOKTRHYFIAAVERLMDYGMSSSPHYLRNRAQSGSVPOFKKVVFOEFTDGSFTOP 871
 Db 1676 GPRSFQOKTRHYFIAAVERLMDYGMST - HYLNRNYQSDNVPQFKKVVFOEFTDGSFSOP 1734
 QY 872 LYRGELNEHGLGPYIRAEVEDNIMVTFERNQASRPYSFYSSLISYEEDQROGAEPKRN 931
 Db 1735 LYRGELNEHGLGPYIRAEVEDNIMVTFERNQASRPYSFYSSLISYKEDOR - GEEPRRN 1793
 QY 932 VKPNETKTYFMKVOHHMARTKDEDFCKAWAFSDVDLEKDVHSGLIGPLVCHTNTLNP 991
 Db 1794 VKPNETKTYFMKVOHHMARTKDEDFCKAWAFSDVDLEKDVHSGLIGPLVCHTNTLNP 991
 QY 992 HGRQVTVQEFALFTTIFDETFKSWYFTENMERNCRAPCNIQMEDPTFKENYRFAINGYIM 1051
 Db 1854 HGRQVTVQEFALFTTIFDETFKSWYFTENMERNCRAPCNIQMEDPTFKENYRFAINGYIM 1051
 QY 1052 DTLPLGLMAQDORIRWYLLSMGSNENIHSIHSGHVFYVRKKEEYKMALYNLYPGVFETV 1111
 Db 1914 DTLPLGLMAQDORIRWYLLSMGSNENIHSIHSGHVFYVRKKEEYKMALYNLYPGVFETV 1111
 QY 1112 EMUPSKAGIWRVECLIGEHLHAGMSTFLVYSNKCQTPLGMA SGHIRDFQITASGOYQOW 1171
 Db 1974 EMUPSKAGIWRVECLIGEHLHAGMSTFLVYSNKCQTPLGMA SGHIRDFQITASGOYQOW 1171
 QY 1172 APKLARLHYSGSINAMSTKEPESWIKVDLAPMIHGIKTQGAROKFSSLYISOITIMYS 1231
 Db 2034 APKLARLHYSGSINAMSTKEPESWIKVDLAPMIHGIKTQGAROKFSSLYISOITIMYS 1231
 QY 1232 LDGKRWQTYRGNSGTGLMVFFGVNDSSGICKHNFNPPIIARYIRLPHTHYSISTLMEI 1291
 Db 2094 LDGKRWQTYRGNSGTGLMVFFGVNDSSGICKHNFNPPIIARYIRLPHTHYSISTLMEI 1291
 QY 1292 MGCILNSCSMPLGMSKASISDAQTASSYFTNMFATWSPSKARLHLQGRSNAMPQVNNP 1351
 Db 2154 MGCILNSCSMPLGMSKASISDAQTASSYFTNMFATWSPSKARLHLQGRSNAMPQVNNP 1351
 QY 1352 KEWLQYDFQTKMKTGVTITQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKYVFOGN 1411
 Db 2214 KEWLQYDFQTKMKTGVTITQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKYVFOGN 1411
 QY 1412 QDSFTPVVNSLDPPLTRYLRIRHPOSVWHQIALRMEVLCGEAQDLY 1457
 Db 2274 QDSFTPVVNSLDPPLTRYLRIRHPOSVWHQIALRMEVLCGEAQDLY 1457

RESULT 3
 KFHU5
 coagulation factor V precursor - human
 N:Alternate names: coagulation labile factor; proaccelerin
 C:Species: Homo sapiens (man)
 C:Date: 19-May-1989 #sequence_revision 02-Jun-1995 #text_change 31-Oct-1997
 C:Accession: A56172; A42344; A28028; A27498; A25897
 R:Crripe, L.D.; Moore, K.D.; Kane, W.H.
 A:Title: Structure of the gene for human coagulation factor V.
 A:Reference number: A42344; MUID:92232668
 A:Accession: A56172
 A:Molecule type: DNA

A:Residues: 1-2224 <CR1>
A:Cross-references: GB:J05368
A:Accession: A42344
A:Molecule type: DNA
A:Residues: 48-58;79-89;120-130;191-201;239-249;313-323;368-378;428-437;461-471;533-542;
2070;2111-2120;2172-2181 <CR2>
R:Jenny, R.J.; Pittman, D.D.; Toole, J.J.; Kriz, R.W.; Aldape, R.A.; Hewick, R.M.; Kaufm
Proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987
A:Title: Complete cDNA and derived amino acid sequence of human factor V.
A:Reference number: A28028; MUID:87260886
A:Accession: A28028
A:Molecule type: mRNA
A:Residues: 1-857,'R',859-864,'R',866-924,'E',926-1763,'I',1765-2212,'T',2214-2224 <JEN>
A:Cross-references: GB:M16967
A:Note: Parts of this sequence, including the amino end of the mature protein, were dete
R:Kane, W.H.; Ichinose, A.; Hagen, F.S.; Davie, E.W.
Biochemistry 26, 6508-6514, 1987
A:Title: Cloning of cDNAs coding for the heavy chain region and connecting region of hum
A:Reference number: A27498; MUID:88107560
A:Accession: A27498
A:Molecule type: mRNA
A:Residues: 1-1284,'I',1286-1600 <KAN>
A:Cross-references: GB:M17785
A:Note: Parts of this sequence were determined by protein sequencing
R:Kane, W.H.; Davie, E.W.
Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986
A:Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor homolog
A:Reference number: A25897; MUID:86313665
A:Accession: A25897
A:Molecule type: mRNA
A:Residues: 1188-1215,1315-2224 <KA2>
A:Cross-references: GB:M14335
A:Note: Parts of this sequence were determined by protein sequencing
R:Keller, F.G.; Ortel, T.L.; Quinn-Allen, M.A.; Kane, W.H.
Biochemistry 34, 4118-4124, 1995
A:Title: Thrombin-catalyzed activation of recombinant human factor V.
A:Reference number: A56139; MUID:95210278
A:Contents: annotation: thrombin cleavage sites
C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C:Genetics:
A:Gene: GDB:F5
A:Cross-references: GDB:119896; OMIM:227400
A:Map position: 1923-1923
A:Introns: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 65
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa prot
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;
1-28/Domain: signal sequence #status predicted <Sig>
29-2224/Product: coagulation factor V #status predicted <Mat>
29-737/Product: coagulation factor Va heavy chain #status experimental <VAH>
F:29-345/Domain: A1 <DA1>
F:33-329/Domain: ferroxidase repeat homology <FO1>
F:346-691/Domain: A2 <DA2>
F:351-684/Domain: ferroxidase repeat homology <FO2>
F:692-1573/Domain: B <DOB>
F:183-1461/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
F:1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
F:1574-1905/Domain: A3 <DA3>
F:1581-1905/Domain: ferroxidase repeat homology <FO3>
F:1667-1765/Region: phospholipid binding #status predicted
F:1906-2064/Domain: C1 <DC1>
F:1906-2064/Domain: I amino-terminal homology <DN1>
F:2065-2224/Domain: C2 <DC2>
F:2065-2221/Domain: discoidin I amino-terminal homology <DN2>
F:51,55,239,297,460,468,554,741,752,760,776,782,821,938,977,1074,1083,1103,1106,1479,149
F:167-193,248-329,500-526,603-684,1725-1751,1907-2061,2066-2221/Disulfide bonds: #status
F:334-335/Cleavage site: Arg-Asn (protein C) #status predicted
F:363,693,1546/Binding site: sulfate (Tyr) (covalent) #status predicted
F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F:382,1338/Binding site: carboxylate (Asn) (covalent) #status absent
F:534-535/Cleavage site: Arg-Gly (protein C) #status predicted

F:737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experiment
F:1046-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experime
F:1573-1574/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match 30.7%; Score 2391.5; DB 1; Length 2224;
Best Local Similarity 26.1%; Pred. No. 1.3e-155;
Matches 593; Conservative 277; Mismatches 483; Indels 923; Gaps 35;

QY 22 RRYLGAVELSDWYQMSDGLGELLVDARFPPRPVPSKFFPNFNSVY-YKKTLEVEFTDHLFNI 80
Db 32 RQFYVAAQGISWSYRPE-----PTNSSLNSTSVTSFKKIYREYEPY-FKK 75
QY 81 AKRPPMGLGPTIQAEVYDVTVITLKNMASHPVSLHAGVSYWKASEGAEYDDQTSQR 140
Db 76 EKPOSTISGLGPTLYAEVGDIIKVFKNKADKPLSIHPQGIKYSKLSGASLYLDHTFPA 135
QY 141 EKEDDKVFPGSHTYWQVLKENGPMASDPLCLTYSYLSHYDLVKDLSGLIGALLVCRE 200
Db 136 EKMDDAVAPRGRETYEWSISDSGPTHDDPCPLTHIYSHENLIEDFNSGLIGPLICK 195
QY 201 GSLAKEKTO-TLHK-FILFAVEDEGKSWHSETKNSLMQDRDAASARAWPMHTVNGYVN 258
Db 196 GLTEGTOKTFDKQIVLFAVFEDESKSWSSSS-----LMYTVNGYVN 239
QY 259 RSLPLGICHRKSVYWHVIGMTPEVHSFLEGHTFLVNRHQAASLEISPTFLTAOTL 318
Db 240 GTPMDITVCADHISWHLGLMSSGPPELFSIHFNQVLEQNHKVASAITLSATSTANMT 299
QY 319 LMDLGQFLFCHISSHQHGMENAYVKVDSCEPEPQLRMKNEEAEDYDDDLTDEMDVVR 378
Db 300 VGPEGKIWISSLTPKHLQAGMQAYIDIKNCCKTRNLKKTIRE----- 343
QY 379 EDDDNPSFQIRSVAKKHPTVWHYIAEEDWDYAPLVAPDDRYSKSOYLNNGPQRI 438
Db 343 -----QRRHMRWEYFIAAEVITWDYAPVIPANMDKKYRSQHLDNFSNQI 387
QY 439 GRKKYKVRMAYTDEFTKREAI--QHESGILPLLYGEVDTLLIFKNQASRPYNIY 495
Db 388 GKHYKVMYTOYEDEF-TKHTVNPNMKEGILGPIIRAQVRDTLKIIVFNMASSRPYSIX 446
QY 496 PHGIT-----DVRPLYSRLPKGVKHLKDPILPGEIFKYKVTVVEDGPTKSDPRLCT 549
Db 447 PHGVTESPYEDEVNSSFT---SGRNNTMIRAVQGETTYTKWNILEFEDEPTENDAOCLT 502
QY 550 RYSSFFVNMERDLASGLIGPLICYKESVDORGNOIMSDKRNVLFSVFENRSWYLTEN 609
Db 503 RPYSDVDIMRDIASGLIGLLICKRSIDRRGIRQRAADIEQOAVFAVFDEKMSWYLEDN 562
QY 610 IQRLPNPAGVQLEDPEFQASNIMHSINGYVDSL-QLSYCLHEVAYWYILSIGAQTDFL 668
Db 563 INKFCENPDEVKRDDEPKFYESINIMSTINGYVESITTLGFCFDDTVQWHFCSVGTONEIL 622
QY 669 SVFFSGYTFKKMYEDTLTLPFSGETVFMSENPGLWILGCHNS----- 715
Db 623 TIHFTGHSFYGKRHEDTLTLPFMRGESVYVTMDNVGTWMLTSMNSSPRSKRLKFRDV 682
QY 715 -----DFNRN----- 720
Db 683 KCIPTDDEDSYEIEFPPESTVATRKMHDRLEPEDESDADYQNRLLAALGIRSFNRS 742
QY 720 -----GM TAL----- 725
Db 743 SLNQEEEFNLTALALENGTEFVSSNTDIYGSNYSPPSNISKFTVNNLAEPQKAPSHQ 802
QY 725 -----LKVSSCDKNTGDIYEDSYED----- 745
Db 803 ATTAGSPLRHLIGKNSVLSSTAHSPPSEDPIDPLQPDVTGIRLLSLGAGEFKSQEH 862
QY 745 ----- 745
Db 863 AKHKGPKVERDOAAKHHRFSNMKLLAHKVGRLSQDTGSPSGMRPWEDLPQDGTGSPSRMR 922

```
QY 745 -----ISAVLLSKNNA----- 756
Db 923 PWKDPDDLKLLKQNSGSKILVGRWHLASEKGYEIIQDDEDTAVNWNWLI SPQNASRAW 982
QY 756 ----- 756
Db 983 GESTPIANKPGKQSGHBPFRVRHKSIOVRQDGGKSRLLKSQFLIKTRKKKKKHTHAP 1042
QY 756 IEPRSF-----SQ 763
Db 1043 LSPRFEHPLRSEAYNPFSEERLKHSLVHKSNETSLPTDLNQTLPMDGWIASLPDHNQ 1102
QY 764 N----- 765
Db 1103 NSSNDTGQASCPGLYQTVPEEHYQTFPIQDPDQMHSTSDPSHRSSPELSEMLEYDRS 1162
QY 765 -----PVLKRH- 772
Db 1163 HKSPFTDISQMSPSSEHEWQTVISPDLSQVTLSPELSQTNLSPDLSHTTLSPELIQRNL 1222
QY 772 ----- 772
Db 1223 SPALGQMPISPDLSHTTTLSPDLSHTTLSLDLSQTNLSPELSQTNLSPALGQMPISPDLSH 1282
QY 772 ----- 772
Db 1283 TTLSDFSQTNLSPELSHMTLSPELSQTNLSPALGQMPISPDLSHTTSLDLSQTNLSPE 1342
QY 772 -----QREIYRTTLOSQDEEI----- 788
Db 1343 LSQTNLSPALGQMPISPDPSHTTTLSDLSQTNLSPELSQTNLSPDLSEMPLEADLSQIPL 1402
QY 788 ----- 788
Db 1403 TPDLQMTLSPDLGETDLSPNFGQMSLSPDLSDTLTLPDLSQISPPDLQ 1462
QY 788 -----DYDDT 792
Db 1463 IFYPSSESSQSLLOEFNESEFPYDGLGQMPSSPTLNDTFLSKFNPLVIGLSKDGIDY 1522
QY 793 ISVENKE-----DFDIYDE-----DENQSP-----RSFQKTRRH 823
Db 1523 IEIPKEEVQSSSEDDYAEIDVYPYDDYKTDVRTNINSSRDPDNIAMWYLRNNGNRNY 1582
QY 824 FIAAVERLWDYGMSSSPHYLRNR--AQSGSVQ--FKKVFEQEFSDGSEFTQPLRYRELN 878
Db 1583 YIAAEISMDY----SEFVQRETDIEDSDDIPEDTYKKVFRKYLDSTFTKDRPREYE 1638
QY 879 EHLGLGPIYIRAEVEDNIMVTFRNQASRPYSFYSLSIYE-----EDQGAEDPRKN 930
Db 1639 EHLGLGPIIRAEVADVIOVRFKNLASRPYSLHAHGLSYEKSSEKTYEDDSPEWEKEDN 1698
QY 931 FVKPNETKTYFWKVQHHMAPTKDEFCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLP 990
Db 1699 AVQPNSSYTYVWHATERSGPESPGSACRAWAYYSAVNPEKDIHSGLIGPLLICQKGLHK 1758
QY 991 AHGROVTVQEFALFTIFDETQSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAIQYI 1050
Db 1759 DSNMPLDMREFVLLFTFDEKKSWSYEEKSSRWSR-----LTSSEMKKSHFHAINGMI 1812
QY 1051 MDTLPLGLVMAQODRIRWYLLSMGSNENIHSIHFSGHYFTVRKKEEYKALYNLYPGVEET 1110
Db 1813 Y-SLPLKMYEQEWRLHLNLNIGSQDIHVHFHGQTLLENGNKQHDGVMPLLPGSFKT 1871
QY 1111 VEMLPKAGIWRVECLIGEHLHAGMSLFLVYSNKCOTPLGMASGHTRDFQITASGQYQ 1170
Db 1872 LEMKASKPGWMLNTEVGENQORAGMQTFLIMDRDCRMPMGLSTGITSDSQIKASEPLGY 1931
QY 1171 WAKPLARLHYSGSINAMSTKE--PFS--WIKVDLLAPMIHIGITOGAROKFSSLYIS 1224
Db 1932 WEPLRLARLNNGGSYNAWSVEKLAAEFASKPWIQVDMQKEVITITGIQTGAKHYLKSCYTT 1991
QY 1225 QFIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSGKIKHNIFNPPIARIYIRLHPHYHSIR 1284
```

```
Db 1992 EFYVAYSSNOINMOIFKGNSTRNMYFNGNSDASTIKENOFPPIVARIYIRISPTRAYNR 2051
QY 1285 STLRLMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMEAT-WSPSKARLHLOGRSNA 1343
Db 2052 PTLRLLELOGCEVNGCSTPLGMENGKTIENKQITASSFKKSWGWDYWEPRARLNAQGRVNA 2111
QY 1344 WRPOVNNPKEWLQVDFQKTMKYTGVTTOGVKSLTSMYVKEFLISSQDGHQWTLFPQNG 1403
Db 2112 WQAKANNKKOWLEIDLKIKKITAIIITQCKSLSEMYVKSytiHYSEQGVEMKPYRLKS 2171
QY 1404 KV--KVFQGNQDSFTPVNLSLDPPLTRYLRIRHPQSWVHQIALRMEVLGCEADLY 1457
Db 2172 SMVDKIFEGNTNTKGHVKNFENPPIISRIFRVLPTKWNQSIARLLELFGC--DIY 2224

RESULT 4
KEBO5
coagulation factor V precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 20-Mar-1998
C:Accession: A42580; A36497
R:Guinto, E.R.; Esmon, C.T.; Mann, K.G.; Macgillivray, R.T.
J. Biol. Chem. 267, 2971-2978, 1992
A:title: The complete cDNA sequence of bovine coagulation factor V.
A:Reference number: A42580; MUID:92147638
A:Accession: A42580
A:Molecule type: mRNA
A:Residues: 1-2211 <GU>
A:Cross-references: GB:M81440; MID:g163037; PID:g163038
A>Note: sequence extracted from NCBI backbone (NCBIN:80774, NCBIPI:80776)
R:Kalafatis, M.; Jenny, R.J.; Mann, K.G.
J. Biol. Chem. 265, 21580-21589, 1990
A:title: Identification and characterization of a phospholipid-binding site of bovine
A:Reference number: A36497; MUID:91072354
A:Accession: A36497
A:Molecule type: protein
A:Residues: 1566-1570,'X',1572-1581,'X',1583-1584;1673-1676,'X',1678-1679,'X',1681,'X'
R:Xue, J.; Kalafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.
Biochemistry 33, 13109-13116, 1994
A:title: Determination of the disulfide bridges in factor Va heavy chain.
A:Reference number: A55979; MUID:95034740
A:Contents: annotation
A>Note: 566-Cys and 617-Cys were shown to have free sulphydryls
C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa p
A:pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasm
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-2211/Product: coagulation factor V #status predicted <MAT>
F:29-741/Product: coagulation factor Va heavy chain #status predicted <VAH>
F:29-345/Domain: A1 <DA1>
F:33-329/Domain: ferroxidase repeat homology <FO1>
F:346-695/Domain: ferroxidase repeat homology <A2 <DA2>
F:351-688/Domain: ferroxidase repeat homology <FO2>
F:696-1564/Domain: B <DOB>
F:1175-1437/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
F:1565-2211/Product: coagulation factor Va light chain #status predicted <VAL>
F:1565-1892/Domain: A3 <DA3>
F:1572-1892/Domain: ferroxidase repeat homology <FO3>
F:1654-1752/Region: phospholipid binding #status predicted
F:1893-2051/Domain: C1 <DC1>
F:1893-2048/Domain: discoidin I amino-terminal homology <DN1>
F:2052-2211/Domain: C2 <DC2>
F:2052-2208/Domain: discoidin I amino-terminal homology <DN2>
F:167-193,248-329,499-525/Disulfide bonds: #status experimental
F:225,239,297,382,460,553,587,745,756,774,780,902,952,964,1044,1053,1062,1071,1078,10
F:324-335/Cleavage site: Arg-Asn (protein C) #status predicted
F:363,697,1537/Binding site: sulfate (Tyr) (covalent) #status predicted
F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F:533-534/Cleavage site: Arg-Gly (protein C) #status predicted
```

F:607-688,1112-1738,1894-2048,2053-2208/Disulfide bonds: #status predicted
F:741-742/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
F:1034-1035/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
F:1564-1565/Cleavage site: Arg-Ser (thrombin) #status experimental

[illegible]

QY	788	-----	788
Db	921	LQKDPYKILNGEMHLVSEKSYEIIQDANENKTVNKLPSNPONDSTRWGENIPFKNSHG	980
QY	788	-----	788
Db	981	KQSGHPFLVTRKRPLODRDRRNSRLKEGLPIRTRKKKEEKPAYHVLSPRSFHLPLR	1040
QY	788	-----	792
Db	1041	GEVNASFSDRRHNSLLHASNETSLSIDLNOTFPSMNLASLAPDHDQTSFNDTTSQT	1100
QY	792	-----TISVENAKKEDFDIYDED-----	814
Db	1101	SSPDLPTVSPPEHYQIFPIODSDPTHSTAPSNSRPDPHTHSTTAPSNSRSPPTQSPQIP	1160
QY	814	-----	814
Db	1161	NYDLNRRAIPTVSOIFPSLELWQTATSLDLSQSPISPDIGOMALSPDGOESLSPDL	1220
QY	814	-----	814
Db	1221	GQTSLSPLDSQESLSPDLGQTALSPDPSQESLSPDLGQTALSPDPSQESLSPDLGQTALS	1280
QY	814	-----	814
Db	1281	PDPGOESLSPDLGQTSLSPLDSQESLSPDLGQTALSPDPSQESLSPDLGQTALSPDPSQE	1340
QY	814	-----	814
Db	1341	SLSPDLGQTSLSPLDGOESLSPDLGQTALSPDPSQESLSPDLGQTSLSPLDGOESLSPDL	1400
QY	814	-----	814
Db	1401	GQTALSPDLDSQESLSPDLGQTPLSPLDLSLSPDLSQLDLKQTSPLDLNQTSHTSSESS	1460
QY	814	-----	814
Db	1461	QSLPLPEFGQTFPNADIGQMPSPDPDSTLNNFTIPEEFNPLVVGLSRDDGDIIEIIPRQ	1520
QY	814	-----RSFQKTRHYFIAVER	830
Db	1521	KEESSEEDYGEFEFAYANDPYQTDLRTDINSSRNPDNIAAWLRSNTGNRKYIYIAAEI	1580
QY	831	LMDYGMSSPHYLNRQAQSGS---VPQ--FKKVVQOEFTDQSFQPLYRGSLNEHLGL	884
Db	1581	SMDYS-----KFVQSDVDVYPEDTVYKKVFRKIYLDSTFTKLDPOGEYEEHLGIL	1631
QY	885	GPYIRAEVEDNINWTFNRNQAASRPYSFYSSLISYE-----EDQROGAERPKNFVKPNE	936
Db	1632	GPVIRAEVDVIOYRFKNLASRPYSLAHGLSIEKSSSEKGTIEDDSPEWEKEDNAIQPNK	1691
QY	937	TKTYFWKVOHHMAPTKDEFDCAKAMAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAGROY	996
Db	1692	TYTYVNHATTRSGPENPGSACRAMAYYSAVNPREDIHSGLIGPLLICRKGLDKETNMPV	1751
QY	997	TVQEFALFTIFDETksWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPG	1056
Db	1752	DMREFULLEFWMFDEKKSWMYDKKPTRSMRRASS-----EYKNSHEFHAINGMIIYN-LPG	1804
QY	1057	LYMAQDQIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKAMALNLYPGVETVEMLPs	1116
Db	1805	LRMYEQEWVRLHLLNLGGSRDIHVYHFGQTLLENGTQOHQCGWPLLPSPGSKTLEMKAS	1864
QY	1117	KAGIMRVECLIGEHLHAGMSTFLPVYSNKCQFPLGASGHIRDFOITASGOYQOWAPKLA	1176
Db	1865	KPGWLLDTEVEGIORAGMOTPLFIVDRECKMBGLSTGLIADSOIQASEFWGWEKPLA	1924
QY	1177	RLHYGSINAW----STK-EPFSWIKVDLLAMIIHGIKTOGAROKFSSLYISOFIIMY	1230
Db	1925	RLNNGSSYNAMIAEKISTEFPNPEMIQVDMQKEVLLTGIOQGAKAHYLKPYTYTTEFCVAY	1984

[illegible]

```

A:Map position: 3q23-3q25
A:Introns: 1006/3; 1061/1
A>Note: the list of introns is incomplete
C:Function:
A:Description: catalyzes the oxidation of free iron(II) to iron(III) coupled with the
A:Note: iron(III), but not iron(II), is the form bound and transported by transferrin
A:Note: other possible functions are amine oxidase activity, copper transport and hom
C:Superfamily: ferroxidase; ferroxidase repeat homology
C:Keywords: acute phase; alternative splicing; copper; duplication; glycoprotein; oxi
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1069/Product: ferroxidase, long form #status predicted <MATL>
F:20-1060/Product: ferroxidase, short form #status experimental <MATS>
F:20-499/Product: ferroxidase 67k chain #status experimental <K67>
F:23-357/Domain: ferroxidase repeat homology <FER1>
F:373-718/Domain: ferroxidase repeat homology <FER2>
F:501-905/Product: ferroxidase 50k chain #status experimental <K50>
F:733-1059/Domain: ferroxidase repeat homology <FER3>
F:907-1065/Product: ferroxidase 19k chain #status experimental <K19>
F:138-397,762/Binding site: cabohydrate (Asn) (covalent) #status experimental
F:174-200,276-357,534-560,637-718,874-900/Disulfide bonds: #status predicted
F:227-588,926/Binding site: cabohydrate (Asn) (covalent) #status absent
F:295-338,343/Binding site: copper (His, Cys, His) (type 1) #status predicted
F:358/Binding site: cabohydrate (Asn) (covalent) (partial) #status experimental
F:656,699,704,709/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted
F:994,1040,1045,1050/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted

Query Match          22.4%; Score 1745; DB 1: Length 1069;
Best Local Similarity 32.5%; Pred. No. 1.1e-11;
Matches 390; Conservative 197; Mismatches 418; Indels 194; Gaps 23;

QY 5 LSTCFELCLRFCEFSATRRYYIGAVELSWDIWQSDLGE---LPVDARPPRPVPSKSPFNT 61
    |||  ::  ::  |||  ::  |||  |||  |||  ::  |||
Db 6 LGIFLFLCSTP-AWAKERHYIIGIETTDWY-ASDHGEKKLISVDTEHSNIYLQNGDRI 63
    |||  ::  ::  |||  ::  |||  |||  |||  ::  |||

QY 62 SVVYKKTLLVEFTDHLFNIAKPPRPWGLLPTIOAEVYDTVITLKNMASHPVSLHVG 121
    :|||  ::  ::  |||  ::  |||  |||  |||  ::  |||  |||  |||  |||  |||
Db 64 GRLYKKALLYQYTDFTFRTTIKRPVWLGFLPILKAETGDKVYVHLKNLASRPTFHSHG 123
    :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 122 VSYWKASEGAEYDDQTSQREKEDDKVPFGSGSHYVWQVLKENGPMASDPLCLTYSLSHV 181
    ::||  |||  |||  |||  ::  |||  |||  |||  ::  |||  |||  |||  |||
Db 124 ITTYKEHGALYPDNTTDFQRADKVPYGEQYTYMLLATEQSPGEGDNCVTRIYHSIH 183
    ::||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 182 DLVKDLSGLIGALLVCREGSLAKEKTQTL-HKFILLFAVEDEGKSWHSE-----TKNS 234
    ||  |||  |||  |||  |||  |||  |||  |||  ::  |||  |||  |||  |||
Db 184 DAPKDIASGLIGLITCKKDSLDEKEKHIDREYVMFSVVDENFSWLEDNIKTYCSBP 243
    ||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 235 LMQDRDASARAWPKMHTVNGVYVNSRLPGLICGRKSVYVHYIGMTTPREHSLFLEGHT 294
    ||  ::  ::  ::  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 244 EKVDKDNEDFQESNRMYSVNGYTFPSLLGLSMCAEDRVKWTYLFMGNEVDVHAAFFHQQA 303
    ||  ::  ::  ::  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 295 FLVRNHRQASLEISPTFTLAQTLLMDLGOFLFCHISSHQHGMAYVYKVDSCPEEPQL 354
    ::||  ::  ::  |||  ::  |||  |||  |||  |||  |||  |||  |||  |||
Db 304 LTNKNYRIDTINLFPATLEFDAYVMAQNPGEWMLSCQNLNHLKAGLQAFQVQEC----- 358
    ::||  ::  ::  |||  ::  |||  |||  |||  |||  |||  |||  |||

QY 355 RMKNNEAEADYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKKHPTKVHVIYIAEEDWDY 414
    ||  ::  ::  |||  ::  |||  |||  |||  |||  |||  |||  |||  |||
Db 358 ---NKSSSKD-----NIRGKHVRH----YIIAAEIIWNY 385
    ||  ::  ::  |||  ::  |||  |||  |||  |||  |||  |||  |||

QY 415 APL-----VLAPDDRSYKSOYLNGPQIRGRKKYKKVREMAVYTDTEF---KTREAIQ 462
    ||  ::  ::  |||  ::  |||  |||  |||  |||  |||  |||  |||  |||
Db 386 AFSGIDIFTRENLTAPGSDS--AVFEQGTTRIGGSYKKLVYREYTDASFTRNKRERGPEE 443
    ||  ::  ::  |||  ::  |||  |||  |||  |||  |||  |||  |||

QY 463 HESGILGPLLYGEVDTLLIIFKNQASRPYNIYPHGI-----TDVRLPY---SRRLP 511
    |||  ::  |||  |||  ::  |||  |||  |||  |||  |||  |||  |||
Db 444 EHLGILGPVIAEAGDITIRVTFHNKCAVPLSTIEPIGVRFNKNNEGITYSPNYPQSRSPV 503
    |||  ::  |||  |||  ::  |||  |||  |||  |||  |||  |||

QY 512 KGVKHLKDPILPLGEIFKKWTVTVEGDPFTKSDPRCLTRYSSFFNMERDLASGLIGPLL 571
    ||  ::  ::  |||  ::  |||  |||  |||  |||  |||  |||  |||
Db 504 PSASH-----VAPTEFTFYEMTVPKVEGPTNADPVCLAKMYISAYDPTKIDIFGLIGPMK 558
    ||  ::  ::  |||  ::  |||  |||  |||  |||  |||  |||

QY 572 ICYKESVDQRGQIMSDKRNVIILFSVPDENRKSWLYTENIQRFLLNPAGVOLDEPDAQSN 631
    ||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```


Db 559 ICKGSLHANGROKQVDKFEYLFPPVDEENESLLEDNIRMFTAPDQVDKEDENFQESN 618
QY 632 IMHSINGVYFDSLO-LSVCLHEVAYWYILSIGAOTDLSVFSGYTEKHKMYEDITLTF 690
Db 619 KMSMNGFMYGNDGLTMCCKSDSVWYILFSAGNEADVHGIFSGNTYLMRGERDTANLF 678
QY 691 PPSGETVMSMENGMLILGCHNSDFRNKMTALLKVSCKDKNTGDIYEDSYEDISAYLL 750
Db 679 POTSLLTHMWPTEGTENVECLTIDHTYTGKMKQKYTVNOCRRQS-----EDS----- 726
UY 751 SKNNAIEPRFSQNPVYLRHQRREITRTTLQSDQDEIDYDDTISVEMKKEDFDIYEDEN 810
Db 726 ----- 726
QY 811 QSPRSFOKTRHYFIAVERLWDYGMSSP-----HVLNRASGSV-----P 853
Db 726 ----TFYLGERTYIAAVEVWDY----SPQREWEKELHLQEQNVSNFLDKGEFYIGS 777
QY 854 QEKVVEQFTDGSFTQPLYRGELNEHLGLPYIRAEVEDNINMTFRNOASRPYSFYS 913
Db 778 KYKKVYRQYTDSTFRVPERKAEHEHLGILPOLHADVGDKVILFKMATRPSIHA- 837
Db 914 LITSEEDQROGAEPRKNFV--PNETKTYFWKYOHMAPTKDEFDCKAWAYFSDVLEK 970
Db 837 -----HGVOTESSTVPTLPGETTLTYWKLIPERSGAGTEDSACIPWAYYSTVDQVK 887
QY 971 DVHSGILGPLVCHTNTLNPAHQVTVQEFALFTTIDETKSWYFTENMERNCAPCNI 1030
Db 888 DLYSGILGPLVCRRLPYLKVFNPRKL--EFALDFLVFDEENESWYLDNIKTYSDPHEKV 945
QY 1031 QMEDPTFENYRHAINGYIMDPLGLVMAQDQIRIWMYLSMGSNENIHSIHSGHETV 1090
Db 946 NKDEEFIESNKMMAINGRMFGNLOGLTMHVGEVWYLMGMNEIDLHTVHFGHSFOY 1005
QY 1091 RKKEEYKMALYNLYPGVETVEMLPKAGIWRVECLIGEHLHAGSTLFLVYSNKCQTP 1149
Db 1006 KHGCVYSSDVDFIPGTQYLTLEMPRTPGIWLHCHVTDHIHAGMETTYTLVQNGEYR 1064

RESULT 6
A35210
ferroxidase (EC 1.16.3.1) precursor - rat
N:Alternate names: ceruloplasmin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jul-1992 #sequence revision 17-Jul-1992 #text_change 13-Mar-1998
C:Accession: A35210; A41753; A29564; S21692
R:Fleming, R.E.; Gitlin, J.D.
J. Biol. Chem. 265, 7701-7707, 1990
A:Title: Primary structure of rat ceruloplasmin and analysis of tissue-specific gene expression
A:Reference number: A35210; MUID:90237081
Accession: A35210
Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1059 <FILE>
A:Cross-references: GB:J05424
R:Fleming, R.E.; Gitlin, J.D.
J. Biol. Chem. 267, 479-486, 1992
A:Title: Structural and functional analysis of the 5'-flanking region of the rat ceruloplasmin gene
A:Reference number: A41753; MUID:92112697
Accession: A41753
A:Molecule type: DNA
A:Residues: 1-48 <FL2>
A>Note: the authors translated the codon GAA for residue 40 as Gly, GAA for residue 41 as Arg, Aldred, A.R.; Grimes, A.; Schreiber, G.; Mercer, J.F.B.
J. Biol. Chem. 262, 2875-2878, 1987
A:Title: Rat ceruloplasmin. Molecular cloning and gene expression in liver, choroid plexus
A:Reference number: A29564; MUID:87137545
Accession: A29564
A:Molecule type: mRNA
A:Residues: 'NSG', 215-216, 'Y', 218, 'FAT', 222, 'F', 224-226, 'E', 228, 'LL', 231, 'D', 233-235, 'RW'
A:Experimental source: liver
A>Note: the authors translated the codon GCG for residue 60 as Gly and GTG for residue 1 as R. Ryan, T.P.; Grover, T.A.; Aust, S.D.

Arch. Biochem. Biophys. 293, 1-8, 1992
A:Title: Rat ceruloplasmin: resistance to proteolysis and kinetic comparison with human
A:Reference number: S21692; MUID:92117681
Accession: S21692
A:Molecule type: protein
A:Residues: 20-29, 'Q', 902-910 <RYA>
C:Superfamily: ferroxidase; ferroxidase repeat homology
C:Keywords: copper; glycoprotein; oxidoreductase; plasma
F:1-19/Domain: signal sequence #status predicted <Sig>
F:20-1059/Product: ferroxidase #status predicted <Mat>
F:23-356/Domain: ferroxidase repeat homology <FER1>
F:372-712/Domain: ferroxidase repeat homology <FER2>
F:727-1053/Domain: ferroxidase repeat homology <FER3>

Query Match 21.8%; Score 1697; DB 2; Length 1059;
Best local similarity 32.5%; Pred. No. 2.1e-108;
Matches 386; Conservative 188; Mismatches 428; Indels 186; Gaps 23;

QY 5 LSTCFPLCLRPFSATRRYYLGAVELSWDYM-Q-SDLGEL-PVDARFPPRPVPRKFPNTS 62
Db 6 LSAFLFL-HSSLAMTREKHYYIGITEAVWDYASGSEKELISVDTEQSNFYLRNPDRI 64
QY 63 VYKKTLEVEFTHLFIKAPRPPMGLGPTIOAEVYDVTYITLKNMASHPVSLHAGV 122
Db 65 RYKKALYSEYTDGTFKTIDKPAWLGLGVPYKAEVGDKVSVAHKNFASRPYTFHAHV 124
QY 123 SYMKASEGAEYDDOTSQREKEDKVPFGSGHYYWQVLKENGPMASDPLCLTYSLSHVD 182
Db 125 TYTKANEGAIPDNTTDFQRADKLFPGQOYLYLRA-NEPSPEGDSNCVTRYHSHVD 183
QY 183 LVYDLSGLIGALLVCREGLAKETQTL-HKPLLLFAVDEGKSWHSETKNSLM----- 237
Db 184 APRDIASGLIGLILCKKSLHKEEENIDQEFVLMFSVDENLSWYLEDNIKTFCESEPE 243
QY 237 -QDRDAASARAWPKMHTVNGVNSLPLGLGCHRSVYWHVIGTTPVHSIFLEGHTF 295
Db 244 KYDKDNEDFQESNMYISINGTFLGSLPLGSMCAEDRVKWLFGMGNEVDVHSALEHGOAL 303
QY 296 LVNHRQASLEISPTFLTAQTLMDLGOFLFCHISSHQHGMAYVYKVDSCPEEPQLR 355
Db 304 TSKNYHTDIINLPATLIDVSMVAQNPGVWMLSCQNLNHLKAGLQAEFQVRDC----- 357
QY 356 MKNNEEAEDYDDLTDSEMDVVRDDNSPFIQIRSVAKKHPTWVHYIAAEEDWDYA 415
Db 357 --NKPS--DDDIQDRHV-----RH-----YYIAAEETIWDYA 385
QY 416 P-----LVLPADDRSYKQYLNNGPQIRGKRYKKVREMAVTDTEF---KTREAI 461
Db 386 PSGTDTFTGENTLSLGSDSRVFEQ---GATRIGSGYKKLVYREYTDSDFTNRKERGPD 441
QY 462 QHESGILPLVGEVDTLLIFKNQASRPYNYPHGITDVR---PLYSRLPKGVKHL 517
Db 442 EEHLGILGPVIMAEVGDIIKRVTFHNKGOPLSIQPMGVFRFTKENEGTYG---PDGRSSK 498
QY 518 KDEPILPGEIIFKRYKWTVEDEGPTKSDPRCLTRYSSFVNMERDLASGLIGPLCYKES 577
Db 499 QASHVAPKETFTYEMTVPKEMGPYADPVLCKMYYSGVDLTKOIFTLGIGPMKCKKS 558
QY 578 VDQGNQIMSDKRNVLFSVFDEKNSWYLTENIQRLPNPAGVLEDEPEFQASNIMHSIN 637
Db 559 LADGROKQVDKFEYLFATVFDEENESLLDDNIRMTTAPENVDKEDDFQESNMHSMN 618
QY 638 GYVFDLSQ-LSVCLHEVAYWYILSIGAOTDLSVFSGYTEKHKMYEDITLTFPSGET 696
Db 619 GFYGNLPLGNMCLGESIWWYILFSAGNEADVHGIFSGNTYLSKGERDTANLEPHKSLT 678
QY 697 VFMSMENPGLWILGCHNSDFRNKMTALLKVSCKDKNTGDIYEDSYEDISAYLLSKNNAI 756
Db 679 LMTPTDEGSFVDECLTIDHTYTGKMKQKYTVNOC-----KGOFEDVTLY----- 723
QY 757 EPRSFSQNPVYLRHQRREITRTTLQSDQDEIDYDDTISVEMKKEDFDIYEDENOSPRSF 816


```

Db 723 ----- 723
QY 817 OKTRHYFIAVERLMDYGMSSSPHYLRNRAQGSV-----POFKKVFQEF 863
      : | :|:|:| | | | : : : | :|
Db 723 -QGEFTYIAAVEVEMDYSRDMEMELHHLQONVSNAFLDKEEFFIGSKYKKVYREF 781
QY 864 TDGSTQPLRGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEEDQO 923
      | : | : : | :|:|:| | | : | : | :|:|:| :|
Db 782 TDSITREQVKRRAEEHGLIGPLIHADYGAKKVVFKNMATRPSIHA-----H 831
QY 924 GAERKNFVK--PNETKTYFWKVOHHMPTKEFDCAMAYFSDVDLEKDVHSGLIGPL 980
      | : : : : | | :|:| :| :| :| :| :| :| :| :| :| :| :| :|
Db 832 GVKIKSSVAPTLPGEVRTYIMQIPERSGAGTEDSPCIPWAYISTYDVRKDLISGLIGPL 891
QY 981 LVC---HTNLPNPAHQVTVQERALLFTIPEDETKSWYFTENNERNCRAPNIOMEDPTF 1037
      :| : : | | :|:| :| :| :| :| :| :| :| :| :| :| :| :|
Db 892 IVCRKSYVKVENPK----KMEFSLLFVFDENESWYLLDNNINTYDPDHEKVNKDNNEF 946
QY 1038 KENTRFHAINGYIMDTLPGLVMAODQIRIRWYLLISGNSNENIHSIHSGHVFTRKKEEYK 1097
      | : :|:|:| :| | | :|:|:|:| :|:|:|:| :| :| :| :| :| :|
Db 947 IESNKMHAINGKMEGNLOGLTIMHVGDENVNMVYAMGNEIDLTVHFHGHSFOYKHRGHS 1006
QY 1098 MALNYLPGVEFEYEMLPKSKAGIWRVECLIGELHLHAGMSTLEFLVYSNK 1145
      : : :| :| :|:| | | : : :|:|:|:| :| :| :| :| :| :| :|
Db 1007 SDVFDEFFPGTYQTELEMFQTPQIGIWLHLHCHVTDIHAGMVTYTVLPNQ 1054

```

```

RESULT 7
M44258
factor VIII-associated gene B hypothetical protein - human
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 31-Oct-1997
C:Accession: A44258
R:Levinson, B.; Kenwright, S.; Game1, P.; Fisher, K.; Gitschler, J.
Genomics 14, 585-589, 1992
A:Title: Evidence for a third transcript from the human factor VIII gene.
A:Reference number: A44258; MUID:93052386
A:Accession: A44258
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-216 <LEV>
A:Cross-references: GB:M90707; NID:g182316; PID:g182317
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
F:1-53/Domain: discoidin I amino-terminal homology #status atypical <DN1>
F:57-210/Domain: discoidin I amino-terminal homology <DN2>

```

Query Match	14.2%;	Score 1104;	DB 2;	Length 216;
Best Local Similarity	100.0%;	Pred. No. 9.3e-69;		
Matches 208;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1250	VEFGNVDSSGIKHNIFNPPIIARIYIRLPHTHYSIRSTLRLMELMGCDLNSCSMPLGMESKA	1309	
Db	9	VEFGNVDSSGIKHNIFNPPIIARIYIRLPHTHYSIRSTLRLMELMGCDLNSCSMPLGMESKA	68	
QY	1310	ISDAQITASSYFTNMFAWTWSPSKARLHLQGRSNAMRPQVNNPKEWLQYDFQKTMKVTGYT	1369	
Db	69	ISDAQITASSYFTNMFAWTWSPSKARLHLQGRSNAMRPQVNNPKEWLQYDFQKTMKVTGYT	128	
QY	1370	TQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKVKVYQGNQDSFTPYVNSLDPLLTR	1429	
Db	129	TQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKVKVYQGNQDSFTPYVNSLDPLLTR	188	
QY	1430	YLRHPOSWVHQAIALRMEVLGCEAQDLY	1457	
Db	189	YLRHPOSWVHQAIALRMEVLGCEAQDLY	216	

```

RESULT      8
JC4915
ags protein precursor - rat
N:Alternate names: O-acetyl-Cd3 ganglioside
C:Species: Rattus norvegicus (Norway rat)

```

C/Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 13-Nov-1998
C/Accession: J04915
R/Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
Biochem. Biophys. Res. Commun. 225, 932-938, 1996
A/Title: Cloning and expression of cDNA for O-acetylation of Gd3 ganglioside.
A/Reference number: J04915; MUID:96374422
A/Accession: J04915
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-427 <OGU>
A/Cross-references: DDBJ:D84068; NID:g1620006; PID:d1012885; PID:g1620007
A/Experimental source: CST cell
C/Comment: This protein is required for the O-acetylation of disialoganglioside sialic
C/Genetics:
A/Gene: ags
C/Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
F,1-21/Domain: signal sequence #status predicted <SIG>
F,28-60/Domain: EGF homology <EG1>
F,68-107/Domain: EGF homology <EG2>
F,110-267/Domain: discoidin I amino-terminal homology <DN1>
F,271-427/Domain: discoidin I amino-terminal homology <DN2>

```

Query Match      8.5%; Score 663; DB 2; Length 427;
Best Local Similarity 40.3%; Pred. No. 5,3e-38;
Matches 142; Conservative 61; Mismatches 123; Indels 26; Gaps 6;

QY 1124 ECLIGELHAGMSTLEFLVYSNK-----COTPLGMASGHRDFOITASGQY-- 1169
      :||: | | | | | : | | | : | | | | | | | | | | | | | | | |
Db 78 KCLVETEDTORG--DIFTEYICQCPVGYSGIHCELGCS TKLGLGGALADSQISASSVYMG 135
      : | | | | | : | | | | | : | | | | | | | | | | | | | | |
QY 1169 ----GQWAPKLARLHYSGSINAW--STKEPFSWIKVDLLAPMIIHGIKTQGARQKFSLLY 1222
      : | | | | | : | | | | | : | | | | | | | | | | | | | | |
Db 136 FMGLQRMGPPELARLYRTGIYNAMWTASYSYDSKPIQYDFLRKMRVSGWMTQGASPRAGAEY 195
      : | | | | | : | | | | | : | | | | | | | | | | | | | | |
QY 1223 ISOFLIMYSLDGKKMOYYRGNSTGTLMWVEFGNVDSGSIKHNIFNPITARIYIRLHPTHYS 1282
      : | | | | | : : : | | | | | : | | | | | : | | | | | :
Db 196 LKTFKAYASLDGRREFEIQDESQTGDKREFMGNQDNNSLKIMFNPFLAEQYIRLYPYVSC 255
      : | | | | | : : : | | | | | : | | | | | : | | | | | :
QY 1283 IRSTLERELMGCDLNSCSMPLGMEKSAISDAQITASSYFT--NMFA-TWSPSKARLHLQ 1339
      || | | | | | : | | | : | | | | | : | | | | | | | | | |
Db 256 RGCTLREFELLGCELHCSCSEPLGKNNTIPDSQITASSSYKTWNLRAFGWYPHLGRLDNQ 315
      : | | | | | : | | | | | : | | | | | : | | | | | : | | | |
QY 1340 RSNAMPQVNNPKEWLQYDFQKTMKVTGVTTOGVKSLTSMYVKEFLISSQDGHQWTLF 1399
      : | | | | | : | | | | | : | | | | | : | | | | | : | | | |
Db 316 KINAMWTAQNSAKEMLOVDLGTQKRVLTGIIITQARDFGHIQYVASRVAHSDDGQWTVY 375
      : | | | | | : | | | | | : | | | | | : | | | | | : | | | |
QY 1400 FONGKVKVFQGNQDSFTPVVNSLDPDLLTRYLRIRHPQSVWHQIALRMEVLGC 1451
      : | | | | | : | | | | | : | | | | | : | | | | | : | | | |
Db 376 EEQGSVKVFQGNLDNNSHKKNIFEKPEMARIYVRVLELWSHNRITLRELLGC 427

```

RESULT 9
A36479
milk fat globule membrane protein - mouse
C/Species: Mus musculus (house mouse)
C/Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 07-Aug-1998
C/Accession: A36479
R/Stdbbs, J.D.; Lekutis, C.; Singer, K.L.; Bul, A.; Yuzuki, D.; Srinivasan, U.; Parry
Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990
A/Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the
A/Reference number: A36479; MWID:91046008
A/Accession: A36479
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-463 <STU>
A/Cross-references: GB:M38337; NID:g199142; PID:g199143
C/Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
C/Keywords: membrane protein
F;28-60/Domain: EGF homology <EGF1>
F;68-107/Domain: EGF homology <EGF2>
F;147-303/Domain: discoidin I amino-terminal homology <DN1>
F;307-463/Domain: discoidin I amino-terminal homology <DN2>

Db 192 EYLKTFKVAISTDGRQFQFIQVAGRS GDKIFIGVNVNSGLKINLEFDTPLETOYRLVPII 251
OY 1281 YSIRSTLRMELMGCDLNSCSPMLGMSKASISDAQITASSYFTN--MFATWSPSKARLHL 1337
Db 252 CHRGTLRFEELGCELNGCTEPLGKDNTIPNKQITASSYKTKWGLSAFSWEPYARLDN 311
OY 1338 OGSNNAWRPQVNNPKEMQLQVDFQKTMKVTGVTQGVKSLTSMYKKEFLISSQDGHWT 1397
Db 312 OGFNNAWTAQTNASAEMLQIDLSGSKRVGTGITQGARDFGHIOYVAAYRAVAGDGVWT 371
OY 1398 LEFONG--KVKVFGQNDQSTFPVNASLDPPLRLRLRIHPQSWHQIALRMEVLGC 1451
Db 372 EYKDPGASESKIFPGNMDNNSHKKNIFETPFQARFVRIQPVAMHNRITLRLVELLGC 427

RESULT 12

Regulation factor VIII - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 31-Dec-1996
C:Accession: A25945
R:Toole, J.J.; Pittman, D.D.; Orr, E.C.; Murtha, P.; Wasley, L.C.; Kaufman, R.J.
Proc. Natl. Acad. Sci. U.S.A. 83, 5939-5942, 1986
A:Title: A large region (approx 95 kDa) of human factor VIII is dispensable for in vitro
A:Reference number: A25945; MUID:86287369
A:Accession: A25945
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-869 <TOO>

Query Match 7.0%; Score 543; DB 2; Length 869;
Best Local Similarity 75.2%; Pred. No. 2.9e-29;
Matches 100; Conservative 14; Mismatches 19; Indels 0; Gaps 0;
OY 765 PVLKRHOREITRTLOSQDEIDYDTISVEMKKEDFDIYEDENQSPRSFQKTKHYF 824
Db 737 PVLRRHQRDISLPTRPQDEEDKMDYDIFSTETKGEFDIYEDENQSPRSFQKTKHYF 796
OY 825 IAAVERLMDYGMSSSPHYLRNRAQSGVYQFKKVVQGFETDGSFTQPLRYGELNHLGL 884
Db 797 IAAVEQLMDYGMSESPRALRNRAQNGEVPRFKKVVFRERADGSFTNPSYRGELNHLGL 856

OY 885 GPYIRAEVEDNIM 897
857 GPYIRAEVEDNIM 869

RESULT 13

A5 antigen precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Mar-1998
C:Accession: JH0466; JQ0948
R:Takagi, S.; Hirata, T.; Agata, K.; Mochii, M.; Eguchi, G.; Fujisawa, H.
Neuron 7, 295-307, 1991
A:Title: The A5 antigen, a candidate for the neuronal recognition molecule, has homology
A:Reference number: JH0466; MUID:91337458
A:Accession: JH0466
A:Molecule type: mRNA
A:Residues: 1-927 <TAK>
A:Cross-references: GB:D10467; GB:D01077; NID:g2222962; PID:d1001730; PID:g2222963
A:Experimental source: tadpole, brain
A:Note: This protein has motifs homologous to complement components C1r and C1s and to C
C:Comment: This protein is a neuronal cell surface molecule involved in the neuronal rec
C:Superfamily: Xenopus A5 antigen; C1r/C1s repeat homology; discoidin I amino-terminal h
C:Keywords: duplication; glycoprotein; transmembrane protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-927/Product: A5 antigen #status predicted <A5A>
F:27-138/Domain: C1r/C1s repeat homology <C1R1>
F:147-262/Domain: C1r/C1s repeat homology <C1R2>
F:274-424/Domain: discoidin I amino-terminal homology <DN1>

F:430-584/Domain: discoidin I amino-terminal homology <DN2>
F:646-812/Domain: MAM homology <MAM>
F:861-883/Domain: transmembrane #status predicted <TM>
F:150,261,300,523,844/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 5.7%; Score 443; DB 2; Length 927;
Best Local Similarity 29.9%; Pred. No. 2.3e-22;
Matches 138; Conservative 75; Mismatches 168; Indels 80; Gaps 19;

OY 1065 IRWYLLSMGSNENIHSHFSGHVFTVRKKEEYKMALYNLY---PGV-----FETVEM- 1114
Db 136 IRYEVFKTGECSRNFTSSNGVIKSPKPEKYPNALLECTYIIIFAPKQEIIVLEFESFELE 195
OY 1114 LPSKA-----GIMRVECLIGENH--HAGMST----- 1138
Db 196 ADSNAPGGQTCRYDWLGIMDGFPGVGPHIGRYCGQNTPGRVRSFTGILSMIFHTDSAIK 255
OY 1138 -----LFLVYSN-----KQOTPLGMASGHIRDFQITASGQYG-QWAPKLARLHYSGSIN 1185
Db 256 EGFFANFSVYQSNTEDEDFQCKEALGMESGEIHFQDISVSQYSMNWSAERSRLNYE--N 313
OY 1186 AMSTKEPF--SWIKVDLAPMIIGHIKITQGA--RQKFSLSYISQFIIMYSLDGKKWQFYR 1241
Db 314 GWTPEGEDTVKEMIQVDLENLRFVSGIGTQGAISKETKKYFKVSKYKVDISSNGEDWITLK 373
OY 1242 -GNSTGLMVFEGNVDSGKIKNIFNPRIARYIRLHPHYHSIRSTLRMELMGCDLNS-- 1299
Db 374 DGNKH---LVFTGNTDATDVYVYRPFPSKPYLTRFVRLRPVYENGISLRFELYGCKITDYP 430
OY 1299 CSMPLGMSKASISDAQITASSYFTNMFAWSPSKARLHLQGRSNAMRPQVNNP--KEWTLQ 1356
Db 431 CSRMIGMVSGLISDSQITASS--QVDRNMYDELARLVTSRSGWALPSPNTHPYTKEWTLQ 487
OY 1357 VDFQKTMKVTGVTQGVKSLTSMYKKEFLISSQDGHQWTLFFQN--GKYKVFQGNQDS 1414
Db 488 IDLAEEKIVRGVLIQGGKHENKVKFMKFKIGYSNNGTEWEMIMDSKKNRPFEGNTNY 547
OY 1415 FFPVNASLDPPLRLRLRIHPQ--SWVHQIALRMEVLGCEAQ 1454
Db 548 DTPELRTF-AHITTGFIIRIIPERASASGIALRLLELIGCEVE 587

RESULT 14

A47285
milk fat globule protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 31-Oct-1997
C:Accession: A47285
R:Iarocca, D.; Peterson, J.A.; Urra, R.; Kuniyoshi, J.; Bistrain, A.M.; Ceriani, R.L
Cancer Res. 51, 4994-4998, 1991
A:Title: A M-r 46,000 human milk fat globule protein that is highly expressed in huma
A:Reference number: A47285
A:Accession: A47285
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-218 <LAR>
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
F:1-56/Domain: discoidin I amino-terminal homology (fragment) <DN1>
F:60-218/Domain: discoidin I amino-terminal homology <DN2>

Query Match 5.4%; Score 424.5; DB 2; Length 218;
Best Local Similarity 40.6%; Pred. No. 4.5e-22;
Matches 84; Conservative 44; Mismatches 72; Indels 7; Gaps 3;

OY 1251 FFGNVDSGKIKNIFNPRIARYIRLHPHYHSIRSTLRMELMGCDLNSCSPMLGMSKAI 1310
Db 13 FVGNNKNAVHVNLFETPVEAQYVRYPTSCHTACTLRFELLGCELNGCANPLGKNNNSI 72
OY 1311 SDAQITASSYF---TNMFATWSPSKARLHLQGRSNAMRPQVNNPKEMQLQVDFQKTKVT 1366
Db 73 PDKQITASSSYKTKWGLHLF-SWNPSYARLDKQGNFNAWVAGSYGNDOWLQVLDGSSKEVT 131

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 1999, 12:00:26 ; Search time 12.92 Seconds
(without alignments)
3187.841 Million cell updates/sec

Title: US=09=001-039a-47
Perfect score: 7797
Sequence: 1 M0ELSTCFCLLRFCSA.....WVH0IALRMEVLGCEADLLY 1457
Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7340	94.1	2351	1	FA8_HUMAN
2	6306	80.9	2133	1	FA8_PIG
3	6261	80.3	2319	1	FA8_MOUSE
4	2390.5	30.7	2224	1	FA5_HUMAN
5	2370	30.4	2211	1	FA5_BOVIN
6	1741	22.3	1065	1	CERU_HUMAN
7	1694	21.7	1059	1	CERU_RAT
8	1651	21.2	1062	1	CERU_MOUSE
9	663	8.5	427	1	MFGM_RAT
10	657	8.4	463	1	MFGM_MOUSE
11	650	8.3	409	1	MFGM_PIG
12	635	8.1	427	1	MFGM_BOVIN
13	588	7.5	387	1	MFGM_HUMAN
14	458.5	5.9	914	1	NRP_CHICK
15	446.5	5.7	923	1	NRP_MOUSE
16	443	5.7	928	1	NRP_XENLA
17	306.5	3.9	3133	1	HMCT_BOMO
18	198.5	2.5	854	1	TRK3_MOUSE
19	192	2.5	913	1	EDD1_HUMAN
20	185.5	2.4	911	1	EDD1_MOUSE
21	185.5	2.4	910	1	EDD1_RAT
22	149.5	1.9	578	1	TRK3_HUMAN
23	148	1.9	578	1	ASO_TOBAC
24	140.5	1.8	622	1	YAK8_SCHPO
25	140.5	1.8	1196	1	BXCN_CLOBO
26	130	1.7	818	1	DAP2_YEAST
27	130	1.7	4196	1	DYHC_SCHPO
28	130	1.7	949	1	HYSB_STRPN
29	128.5	1.6	994	1	DPO2_KLULA
30	125	1.6	2167	1	BEM2_YEAST
31	125	1.6	2368	1	ESR1_YEAST
32	123	1.6	587	1	ASO_CUCSA
33	123	1.6	1471	1	MYB4_YEAST
34	123	1.6	867	1	SSPO_BOVIN
35	121.5	1.6	2216	1	YCF2_EPIVI
36	121.5	1.6	2280	1	YCF2_TOBAC
37	120.5	1.5	1405	1	DPOA_SCHPO
38	119.5	1.5	1098	1	RPOP_MAIZE
39	119.5	1.5	1648	1	YJ9H_YEAST
40	119	1.5	1165	1	LEPR_HUMAN
41	118.5	1.5	923	1	PWP2_YEAST
42	118.5	1.5	2167	1	YCS2_YEAST
43	117.5	1.5	1430	1	GTFD_STRMU

ALIGNMENTS

RESULT	1	44	117.5	1.5	1139	1	HMW1_MYCGE
FA8_HUMAN		45	117.5	1.5	1461	1	P87078 candida alb
ID	FA8_HUMAN						
AC	P00451;						
DT	21-JUL-1986 (REL. 01, CREATED)						
DT	21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)						
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)						
DE	COAGULATION FACTOR VIII PRECURSOR (PROCOAGULANT COMPONENT).						
GN	F8C.						
OS	HOMO SAPIENS (HUMAN).						
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;						
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE; 86081164.						
RA	TRUETT M.A., BLACHER R., BURKE R.L., CAPUT D., CHU C., DINA D.,						
RA	HARTOG R., KUO C.H., MASIAZ F.R., MERRYWEATHER J.P., NAJARIAN R.,						
RA	PACHL C., POTTER S.J., PUMA J., QUIROGA M., RALL L.B., RANDOLPH A.,						
RA	URDEA M.S., VALENZUELA P., DAHL H.-H.M., FAVALLARO J., HANSEN J.,						
RA	NORDEANG O., EZBAN M.;						
RT	"Characterization of the polypeptide composition of human factor						
RT	VIII:C and the nucleotide sequence and expression of the human kidney						
RT	CDNA.";						
RL	DNA 4:333-349(1985).						
RN	[2]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE; 85061548.						
RA	WOOD W.I., CAPON D.J., SIMONSEN C.C., EATON D.L., GITSCHIER J.,						
RA	KEYT B., SEEBURG P.H., SMITH D.H., HOLLINGSHEAD P., WION K.L.,						
RA	DELMART E., TUDDEHAM E.G.D., VEHAR G.A., LAWN R.M.;						
RT	"Expression of active human factor VIII from recombinant DNA clones.";						
RL	NATURE 312:330-337(1984).						
RN	[3]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE; 85061550.						
RA	TOOLE J.J., KNOPE J.L., WOZNEY J.M., SULTZMAN L.A., BUECKER J.L.,						
RA	PITTMAN D.D., KAUFMAN R.J., BROWN E., SHOEMAKER C., ORR E.C.,						
RA	AMPHLETT G.W., FOSTER W.B., COE M.L., KNUTSON G.J., FASS D.N.,						
RT	HEWICK R.M.;						
RT	"Molecular cloning of a cDNA encoding human antihemophilic factor.";						
RL	NATURE 312:342-347(1984).						
RN	[4]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE; 93265012.						
RA	GITSCHIER J., WOOD W.I.;						
RT	"Sequence of the exon-containing regions of the human factor VIII						
RT	gene.";						
RL	HUM. MOL. GENET. 1:199-200(1992).						
RN	[5]						
RP	SEQUENCE OF 2064-2070 FROM N.A.						
RA	DE WATER N.S., WILLIAMS R., BROWETT P.J.;						
RL	SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.						
RN	[6]						
RP	SULFATATION OF 1699.						
RX	MEDLINE; 91093266.						
RA	LEYTE A., VAN SCHIJNDEL H.B., NIEHRS C., HUTTNER W.B., VERBEET M.P.,						
RA	MERTENS K., VAN MOERIK J.A.;						
RT	"Sulfation of Tyrl680 of human blood coagulation factor VIII is						
RT	essential for the interaction of factor VIII with von Willebrand						
RT	factor.";						
RL	J. BIOL. CHEM. 266:740-746(1991).						
RN	[7]						
RP	SULFATATION.						
RX	MEDLINE; 92207952.						
RA	PITTMAN D.D., WANG J.H., KAUFMAN R.J.;						
RT	"Identification and functional importance of tyrosine sulfate						

RT residues within recombinant factor VIII.";
RL BIOCHEMISTRY 31:3315-3325(1992).
RN [8]
RP STRUCTURE BY NMR OF 2322-2343.
RX MEDLINE; 95200924.
RA GILBERT G.E., BALEJA J.D.;
RT "membrane-binding peptide from the C2 domain of factor VIII forms an
RL amphipathic structure as determined by NMR spectroscopy.";
RN BIOCHEMISTRY 34:3022-3031(1995).
RN [9]
RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE; 91221499.
RA GITSCHIER J.;
RT "The molecular basis of hemophilia A.";
RN ANN. N.Y. ACAD. SCI. 614:89-96(1991).
RN [10]
RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE; 89088506.
RA WHITE G.C. II, SHOEMAKER C.B.;
RT "Factor VIII gene and hemophilia A.";
RN BLOOD 73:1-12(1989).
RL [11]
RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE; 95245332.
RA ANTONARAKIS S.E., KAZAZIAN H.H., TUDDENHAM E.G.D.;
RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
RN HUM. MUTAT. 5:1-22(1995).
RN [12]
RP VARIANT GLN-2326.
RX MEDLINE; 86235434.
RA GITSCHIER J., WOOD W.I., SHUMAN M.A., LAMN R.M.;
RT "Identification of a missense mutation in the factor VIII gene of a
RL mild hemophilic.";
RN SCIENCE 232:1415-1416(1986).
RN [13]
RP VARIANT PRO-2135.
RX MEDLINE; 88096539.
RA LEVINSON B., JANCO R.L., PHILLIPS J.A. III, GITSCHIER J.;
RT "A novel missense mutation in the factor VIII gene identified by
RL analysis of amplified hemophilia DNA sequences.";
RN NUCLEIC ACIDS RES. 15:9797-9805(1987).
RN [14]
RP VARIANT GLN-2228.
RX MEDLINE; 88191889.
RA YOUSSEFIAN H., ANTONARAKIS S.E., BELL W., GRIFFIN A.M.,
RN KAZAZIAN H.H.;
RT "Nonsense and missense mutations in hemophilia A: estimate of the
RL relative mutation rate at CG dinucleotides.";
RN AM. J. HUM. GENET. 42:718-725(1988).
RN [15]
RP VARIANT GLY-291.
RX MEDLINE; 88220354.
RA YOUSSEFIAN H., WONG C., ARONIS S., PLATOKOUKIS H., KAZAZIAN H.H. JR.,
RN ANTONARAKIS S.E.;
RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
RL in exon 7 of the factor VIII gene.";
RN AM. J. HUM. GENET. 42:867-871(1988).
RN [16]
RP VARIANT CYS-1708.
RX MEDLINE; 89274393.
RA O'BRIEN D.P., TUDDENHAM E.G.;
RT "Purification and characterization of factor VIII 1,689-Cys: a
RN nonfunctional cofactor occurring in a patient with severe hemophilia
RL A.";
RN BLOOD 73:2117-2122(1989).
RN [17]
RP VARIANT CYS-391.
RX MEDLINE; 90001543.
RA SHIMA M., WARE J., YOSHIOKA A., FUKUI H., FULCHER C.A.;
RT "An arginine to cysteine amino acid substitution at a critical
RN thrombin cleavage site in a dysfunctional factor VIII molecule.";
RL BLOOD 74:1612-1617(1989).
RN [18]

RP VARIANT LEU-189.
RX MEDLINE; 90057680.
RA CHAN V., CHAN T.K., TONG T.M., TODD D.;
RT "A novel missense mutation in exon 4 of the factor VIII:C gene
RL resulting in moderately severe hemophilia A.";
RN BLOOD 74:2688-2691(1989).
RP VARIANT LEU-2326.
RX MEDLINE; 89197216.
RA INABA H., FUJIMAKI M., KAZAZIAN H.H. JR., ANTONARAKIS S.E.;
RT "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26
RL of the factor VIII gene.";
RN HUM. GENET. 81:335-338(1989).
RP VARIANT HIS-391.
RX MEDLINE; 89264602.
RA ARAI M., INABA H., HIGUCHI M., ANTONARAKIS S.E., KAZAZIAN H.H. JR.,
RA FUJIMAKI M., HOYER L.W.;
RT "Direct characterization of factor VIII in plasma: detection of a
RL mutation altering a thrombin cleavage site
[arginine-372-->histidine]";
RN PROC. NATL. ACAD. SCI. U.S.A. 86:4277-4281(1989).
RP VARIANT CYS-1708.
RX MEDLINE; 90105723.
RA ARAI M., HIGUCHI M., ANTONARAKIS S.E., KAZAZIAN H.H. JR.,
RA PHILLIPS J.A. III, JANCO R.L., HOYER L.W.;
RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
RL Cys) in the factor VIII gene of two unrelated patients with
cross-reacting material-positive hemophilia A.";
RN BLOOD 75:384-389(1990).
RP [22]
RX VARIANTS GLN-2228 AND LEU-2326.
RA MEDLINE; 90123183.
RA CASULA L., MURRU S., PECORARA M., RISTALDI M.S., RESTAGNO G.,
RA MANCUSO G., MORFINI M., DE BIASI R., BAUDO F., CARBONARA A.;
RT "Recurrent mutations and three novel rearrangements in the factor
RL VIII gene of hemophilia A patients of Italian descent.";
RN BLOOD 75:662-670(1990).
RP [23]
RX VARIANT CYS-391.
RA MEDLINE; 90329422.
RA PATTINSON J.K., MCVEY J.H., BOON M., AJANI A., TUDDENHAM E.G.;
RT "C6M+ haemophilia A due to a missense mutation (372-->Cys) at the
RL internal heavy chain thrombin cleavage site.";
RN BR. J. HAEMATOL. 75:73-77(1990).
RP [24]
RX VARIANTS PHE-1699 AND CYS-1708.
RA MEDLINE; 90152691.
RA HIGUCHI M., WONG C., KOCHHAN L., OLEK K., ARONIS S., KASPER C.K.,
RA KAZAZIAN H.H., ANTONARAKIS S.E.;
RT "Characterization of mutations in the factor VIII gene by direct
RL sequencing of amplified genomic DNA.";
RN GENOMICS 6:65-71(1990).
RP [25]
RX VARIANTS CYS-1728 AND ASP-1941.
RA MEDLINE; 90169988.
RA TRAYSTMAN M.D., HIGUCHI M., KASPER C.K., ANTONARAKIS S.E.,
RA KAZAZIAN H.H.;
RT "Use of denaturing gradient gel electrophoresis to detect point
RL mutations in the factor VIII gene.";
RN GENOMICS 6:293-301(1990).
RP [26]

[illegible]

Db 61 TSYYKKTLEFVEFDHLENIAPRPPWMLGPTIOAEVYDTVITLKNMASHPSVLHAY 120
QY 121 GVSYWKASEGAEDDQTSOREKEDDKYFPGSHYVQVLKENGPMASDPLCLTYSYLH 180
Db 121 GVSYWKASEGAEDDQTSOREKEDDKYFPGSHYVQVLKENGPMASDPLCLTYSYLH 180
QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTOTLHKFILLFAVFEDEGKSWHSETKNSLMODRD 240
Db 181 VDLVKDLNSGLIGALLVCREGSLAKEKTOTLHKFILLFAVFEDEGKSWHSETKNSLMODRD 240
QY 241 AASARAWPKMHTVNGVYVNRSLPGLIGCHRSYVYHIGMGTTPVHSHIFLEGHTFLVRNH 300
Db 241 AASARAWPKMHTVNGVYVNRSLPGLIGCHRSYVYHIGMGTTPVHSHIFLEGHTFLVRNH 300
QY 301 RQASLEISPTITFLAOTLLMDLGQFLLFCHISSHQHDMGMEAYVYKVDSCPEEPQLMKKNE 360
Db 301 RQASLEISPTITFLAOTLLMDLGQFLLFCHISSHQHDMGMEAYVYKVDSCPEEPQLMKKNE 360
QY 361 EAEDYDDDLTDEMDVYRFDDDNSPSFIQIRSVAKKHPTWVHYIAAEEDWDYADLYLA 420
Db 361 EAEDYDDDLTDEMDVYRFDDDNSPSFIQIRSVAKKHPTWVHYIAAEEDWDYADLYLA 420
QY 421 PDDSYKSOYLNNGPORIGRKYKRYPMAYTDETFKTREAIQHESGILGPLLYGEVDTL 480
Db 421 PDDSYKSOYLNNGPORIGRKYKRYPMAYTDETFKTREAIQHESGILGPLLYGEVDTL 480
QY 481 LIIRKNOASRPYNIYPHGITDVRLPYSRLLPKGVKHLKDFPILPGEIFKYKWTYVEDGP 540
Db 481 LIIRKNOASRPYNIYPHGITDVRLPYSRLLPKGVKHLKDFPILPGEIFKYKWTYVEDGP 540
QY 541 TKSDBRCLTRYSSSEYVNMERDLASGLIGPLLCYKESVDQRGNOIMSDKRNVLFSYFDE 600
Db 541 TKSDBRCLTRYSSSEYVNMERDLASGLIGPLLCYKESVDQRGNOIMSDKRNVLFSYFDE 600
QY 601 NRSWYLTENIQRLPNPAGVOLLEDPFOASNIMHSINGYVDSLOISVCLHEVAYWYILS 660
Db 601 NRSWYLTENIQRLPNPAGVOLLEDPFOASNIMHSINGYVDSLOISVCLHEVAYWYILS 660
QY 661 IGAQTDFLSVFSGYTFKHKMYEDTLTLPFSGEYVFMSEMPGLMILGCHNSDFPRNG 720
Db 661 IGAQTDFLSVFSGYTFKHKMYEDTLTLPFSGEYVFMSEMPGLMILGCHNSDFPRNG 720
QY 721 MTALIKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSF----- 762
Db 721 MTALIKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSF----- 762
QY 762 ----- 762
Db 762 ----- 762
QY 841 PGALDSNNSLSEMTHERPOLHSGDMYFTPESGLOLRNLKLTATATELKLDFKYSST 900
Db 841 PGALDSNNSLSEMTHERPOLHSGDMYFTPESGLOLRNLKLTATATELKLDFKYSST 900
QY 901 SNNLISTIPSDNLAAGTDNTSSLGPPMPVHYDSQDLTTLFGKSSPLTESGGLSSEE 960
Db 901 SNNLISTIPSDNLAAGTDNTSSLGPPMPVHYDSQDLTTLFGKSSPLTESGGLSSEE 960
QY 961 NNDKLESLGLMNSQESSWGKNVSTESGRLLFKGRAHGPAALLTKNALFKVISLTKTN 1020
Db 961 NNDKLESLGLMNSQESSWGKNVSTESGRLLFKGRAHGPAALLTKNALFKVISLTKTN 1020
QY 1021 KTSNNSATNRKTHIDGSLLIENSFSWQNILSDTEFKKVTPLIHDMMLMDKNATLRL 1080
Db 1021 KTSNNSATNRKTHIDGSLLIENSFSWQNILSDTEFKKVTPLIHDMMLMDKNATLRL 1080
QY 1081 NHMSNKTTSKNMEMVQOKKEGPIPPDAQNPDMSFFKMLFLPESARWIORTHGKNSLNSG 1140
Db 1081 NHMSNKTTSKNMEMVQOKKEGPIPPDAQNPDMSFFKMLFLPESARWIORTHGKNSLNSG 1140
QY 762 ----- 762

Db 1141 OGPSPKOLVSLGPEKSVEGQNFLESEKNKVVGGEFTKDVGLKEMVFPSSRNLEFLTNLDN 1200
QY 762 ----- 762
Db 1201 LHENNTNQEKIQQEIEKKEKELLIOENVLPOIHVTGTCKNFMKNLFLSTROVEGSYD 1260
QY 762 ----- 762
Db 1261 GAYAPVLQDFRSLNDSTNRTKKHTAHFSKKEEENLEGLNQOTQOIVEKYACETRISPNT 1320
QY 762 ----- 762
Db 1321 SQONFVTQSKKRALKQFLRPLEETELEKRRIYDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
QY 762 ----- 762
Db 1381 KGATQSPSLSDCLTRSHSIPQANSPLPIAKVSSFPSIRPIYTLRVLFQDNSSHLPAASY 1440
QY 762 ----- 762
Db 1441 RKKDSGOESSHFLQAKKNNLSAILLTEMGTQREVSGISGTSATNSVTYKKVENTVLP 1500
QY 762 ----- 762
Db 1501 KPDLPTSGKVLLPKVHIYQKLEPTETSNPSGHLDLVEGSLLOGTEGAIKWNEANRP 1560
QY 762 ----- 762
Db 1561 GKVPFLRVATESSAKTPSKLLDPLAMDNHYGTQIPKEEMKSOEKSDEKTAFFKKKDTILSL 1620
QY 762 ----- 762
Db 1621 NACESNHAIAINEGQNKPEIEYTWAKQGRTERLCSQNPVLKRHOREITRTTLOSQOE 1680
QY 762 ----- 762
Db 1681 IDYDITISVEMKKEDFDIYDEDENSPRSFOKTRHYFIAAVERLMDYGMSSSPHYLRNR 1740
QY 847 AOGSVPOFKKVVQEFDTGSGFTQPLRGELNEHGLGPYIRAEYEDNIMVTFRNOASR 906
Db 1741 AOGSVPOFKKVVQEFDTGSGFTQPLRGELNEHGLGPYIRAEYEDNIMVTFRNOASR 1800
QY 907 PYSFTSSLISYEEDOROGAEPKRNKVPNETKTYFWKVQHHMAPTDEFDCKAMAFSDV 966
Db 1801 PYSFTSSLISYEEDOROGAEPKRNKVPNETKTYFWKVQHHMAPTDEFDCKAMAFSDV 1860
QY 967 DLEKDVHSGLIGPLVCHTNTLNPAHGRQVTVQEFALFTTIFDETKSWYFTENMERNCR 1026
Db 1861 DLEKDVHSGLIGPLVCHTNTLNPAHGRQVTVQEFALFTTIFDETKSWYFTENMERNCR 1920
QY 1027 PCNIQMEDPTFKENYRFHAINGIYMDTLPLGLVMAQDQIRIRWYLLSGSNSNENIHSIHFSGH 1086
Db 1921 PCNIQMEDPTFKENYRFHAINGIYMDTLPLGLVMAQDQIRIRWYLLSGSNSNENIHSIHFSGH 1980
QY 1087 VFTYRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 1146
Db 1981 VFTYRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
QY 1147 QTPLGMAAGHIRDQITASGOYGOWAPKLARLHYSGSINAWSTKEBFSWIKVDLAPMI 1206
Db 2041 QTPLGMAAGHIRDQITASGOYGOWAPKLARLHYSGSINAWSTKEBFSWIKVDLAPMI 2100
QY 1207 HGIRTOGAROKFSSSYISOFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSGIRKNIEN 1266
Db 2101 HGIRTOGAROKFSSSYISOFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSGIRKNIEN 2160
QY 1267 PRIARVIRLPHPTHSIRSTLRMELMGCDLNSCMBPLGEMSKAISDAQITASSYFTNMFA 1326
Db 2161 PRIARVIRLPHPTHSIRSTLRMELMGCDLNSCMBPLGEMSKAISDAQITASSYFTNMFA 2220
QY 1327 TWSPSKARLHLQGRSNAMRPQVNNPKEWLQVDFOKTMKVTGVTTOGVKSLLTSMYKEFL 1386
Db 2221 TWSPSKARLHLQGRSNAMRPQVNNPKEWLQVDFOKTMKVTGVTTOGVKSLLTSMYKEFL 2280

```

QY      1387 ISSSDGHWTLTFPONGKVKVFOGNDSFTPVNSLDPLLTRYLRIHPQSWHQIALRM 1446
          |||
Db      2281 ISSSDGHQWTLFPONGKVKVFOGNDSFTPVNSLDPPLLTRILRIHPQSWHQIALRM 2340
          |||
QY      1447 EVLGCEAODLY 1457
          |||
Db      2341 EVLGCEAODLY 2351

RESULT      2
FAB_PIG
ID_FAB_PIG STANDARD; PRT; 2133 AA.
AC P12263; Q95243;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
COAGULATION FACTOR VIII PRECURSOR (PROCOAGULANT COMPONENT).
CF8.
SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RA SEQUENCE FROM N.A.
RA HEALEY J.F., LUBIN I.M., LOLLAR P.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RA SEQUENCE OF 705-1573 FROM N.A.
RX MEDLINE; 86287369.
RA TOOLE J.J., PITTMAN D.D., ORR E.C., MURTHA P., WASLEY L.C.,
RA KAUFMAN R.J.;
RT "A large region (approximately equal to 95 kDa) of human factor VIII
RT is dispensable for in vitro procoagulant activity.";
RL PROC. NATL. ACAD. SCI. U.S.A. 83:5939-5942(1986).
RN [3]
RA SEQUENCE OF 392-759 FROM N.A.
RX MEDLINE; 94179260.
RA LUBIN I.M., HEALEY J.F., SCANDELLA D., RUNGE M.S., LOLLAR P.;
RT "Elimination of a major inhibitor epitope in factor VIII.";
RL J. BIOL. CHEM. 269:8639-8641(1994).
CC -I- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS
CC AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE
CC ACTIVATED FORM, FACTOR XA.
CC -I- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -I- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -I- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -I- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC -----
DR EMBL; U49517; G1511634; -.
DR PIR; A25945; A25945.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PFAM; PF00394; Cu-oxidase; 3.
DR PFAM; PF00754; F5_F8_type_C; 2.
DR HSSP; P00451; ICFG.
KM BLOOD COAGULATION; REPEAT; PLASMA; ACUTE PHASE; CALCIUM;
KW SIGNAL; GLYCOPROTEIN; SULFATATION.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2133 COAGULATION FACTOR VIII.
FT DOMAIN 20 357 F5/8 TYPE A 1.
FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.
FT DOMAIN 207 357 PLASTOCYANIN-LIKE 2.
FT DOMAIN 399 730 F5/8 TYPE A 2.
```

FT	DOMAIN	399	573	PLASTOCYANIN- LIKE 3.
FT	DOMAIN	583	730	PLASTOCYANIN- LIKE 4.
FT	DOMAIN	760	1599	B.
FT	DOMAIN	1495	1822	F5/8 TYPE A 3.
FT	DOMAIN	1495	1659	PLASTOCYANIN- LIKE 5.
FT	DOMAIN	1669	1822	PLASTOCYANIN- LIKE 6.
FT	DOMAIN	1822	1970	F5/8 TYPE C 1.
FT	DOMAIN	1975	2127	F5/8 TYPE C 2.
FT	SITE	391	392	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT	SITE	759	760	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT	SITE	1449	1450	CLEAVAGE (ACTIVATION) (BY SIMILARITY).
FT	SITE	1490	1491	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT	MOD_RES	737	737	SULFATATION (BY SIMILARITY).
FT	MOD_RES	738	738	SULFATATION (BY SIMILARITY).
FT	MOD_RES	742	742	SULFATATION (BY SIMILARITY).
FT	DISULFID	173	199	PROBABLE.
FT	DISULFID	547	573	PROBABLE.
FT	DISULFID	1633	1659	PROBABLE.
FT	DISULFID	1822	1970	BY SIMILARITY.
FT	DISULFID	1975	2127	BY SIMILARITY.
FT	CARBOHYD	233	233	POTENTIAL.
FT	CARBOHYD	259	259	POTENTIAL.
FT	CARBOHYD	601	601	POTENTIAL.
FT	CARBOHYD	929	929	POTENTIAL.
FT	CARBOHYD	985	985	POTENTIAL.
FT	CARBOHYD	1025	1025	POTENTIAL.
FT	CARBOHYD	1111	1111	POTENTIAL.
FT	CARBOHYD	1181	1181	POTENTIAL.
FT	CARBOHYD	1208	1208	POTENTIAL.
FT	CARBOHYD	1245	1245	POTENTIAL.
FT	CARBOHYD	1265	1265	POTENTIAL.
FT	CARBOHYD	1335	1335	POTENTIAL.
FT	CARBOHYD	1408	1408	POTENTIAL.
FT	CARBOHYD	1611	1611	POTENTIAL.
FT	CARBOHYD	1919	1919	POTENTIAL.
FT	CONFLICT	713	713	N -> M (IN REF. 2).
FT	CONFLICT	734	734	I -> T (IN REF. 2).
FT	CONFLICT	792	792	G -> Q (IN REF. 2).
FT	CONFLICT	1133	1133	E -> F (IN REF. 2).
FT	CONFLICT	1191	1191	I -> L (IN REF. 2).
FT	CONFLICT	1209	1209	R -> F (IN REF. 2).
FT	CONFLICT	1437	1437	C -> G (IN REF. 2).
FT	CONFLICT	1456	1456	F -> R (IN REF. 2).
FT	CONFLICT	1539	1539	F -> R (IN REF. 2).
FT	CONFLICT	1546	1546	Q -> N (IN REF. 2).
SQ	SEQUENCE	2133	AA: 239304	MM: 07ECA01E CRC32: 239304

Query Match	80.9%	Score 6306	DB 1	Length 2133
Best Local Similarity	57.5%	Pred. No. 0		
Matches 1227	Conservative 107	Mismatches 122	Indels 678	Gaps 3
Qy	1 MÖEISTCFELCLLRFCSATRRYYLGAVELSWDYMOSD-LGELPYDARFPPRVKSFPF	59		
	: : :			
Db	1 MÖEISTCVFLCLLPLGFSAIRRYYLGAVELSWDYRQSELLRELHVDTRFPATAPGALPL	60		
Qy	60 NTSVYVKKTLFEVETDHLFNIAKPRPPMGLGPITQAEVYDVIVITLKNMASHPVSLHA	119		
	: : : : :			
Db	61 GPSVLTAKTVFEVETDQLFSVARPRPPMGLGPITQAEVYDVIVITLKNMASHPVSLHA	120		
Qy	120 VGVSTWKASEGAEXDDQTSQREKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTYSYLS	179		
	: :			
Db	121 VGVSEWKSSEGAEXEDHTSQREKEDDKVLPGKSQTYWQVLKENGFTASDPCLTYSYLS	180		
Qy	180 HVDLYKDLNSGLIGALLVCREGSLAKEKTQTLHKEFLLFVAFDEGKSWHSETKNSLMQDR	239		
Db	181 HVDLYKDLNSGLIGALLVCREGSLTRERTQNLHEFVLLFAVFDEGKSWHSARNDSTWTAM	240		
Qy	240 DAASARAWPKMHTVNGVYVNRSLPGLIGCHRSVYWHVIGMTTPEVHSIFLEGHTFLVRN	299		
Db	241 DPAPARAQPMHTVNGVYVNRSLPGLIGCHKKSVMYHVIIGMTSPEVHSIFLEGHTFLVRH	300		
Qy	300 HRQASLEISPTITTAQTLMLMDIGQFLFCHISSHOHDGMEAYKVDSCPEEPQLRMKNN	359		

Db 301 HRQASLEISPLFTLAQTFLMDIGQFLLFCHISSHHGMEAHVRESCAEEDPOLRRKAD 360
QY 360 EEAEDYDDDLTJSEMDVVRFDDBNSPSFQIRSAVAKKHPKTWVHYIAAEEEDMDYAPLV 419
Db 361 EE-EDYDNLVSDMDVRLDGDVSPFIQIRSAVAKKHPKTWVHYISAEEDMDYAPAVP 419
QY 420 APDRSRSQYLNNGPQRIGRKKYKVRFMAYTDETFKTRERAIQHESGILGPLLXGEVGT 479
Db 420 SPDSRSYKSLYNSGPQRIGRKKYKARFVAYTDTYFKTRKAIPRESGILGPLLXGEVGT 479
QY 480 LLIIFKNQASRPYNYPHGITDVRPLYSRRLPKGVKHLKDFPLPGEIFKXKWTYVEDG 539
Db 480 LLIIFKNKASRPYNYPHGITDYSALHPGRLLKGWKLKMDPILPGETFKYKWTYVEDG 539
QY 540 PTKSDPRLCTRYSSFVNMERDLASGLIGPLICYKESVDQRCNOIMSDKRNVLFSVD 599
Db 540 PTKSDPRLCTRYSSSINLEKDLASGLIGPLICYKESVDQRCNOIMSDKRNVLFSVD 599
QY 600 ENRSWYLTENIORFLPNPAGVQLEDEPEFQASINMHSINGYVFDLSQLSVCLHEVAYWYL 659
Db 600 ENOSWYLAENIORFLPNPDGLQDPODEFOASINMHSINGYVFDLSQLSVCLHEVAYWYL 659
QY 660 SIGAQTDFLSVFFSGYTFKHKNYEDTLTFPFSGETVFMSEMPGLMILGCHNSDFRNR 719
Db 660 SYGAQTDFLSVFFSGYTFKHKNYEDTLTFPFSGETVFMSEMPGLMILGCHNSDLNR 719
QY 720 GMTALLKVSSCKNTGDYEDSDISAYLLSKNNAIEPRFSQN----- 765
Db 720 GMTALLKVSSCKNTGDYEDSDISAYLLSKNNAIEPRFSQN----- 765
QY 780 SPEEDVELDPQSGERTQALEELSVPSGDGSMILGONPAPHGSSSSDLQEARNEADYLP 839
Db 780 SPEEDVELDPQSGERTQALEELSVPSGDGSMILGONPAPHGSSSSDLQEARNEADYLP 839
QY 840 ABERNTAPSAARLRLPELHSAERYLTPEPEKELKLDKSSSSDLKTSPTIPSDTLS 899
Db 840 ABERNTAPSAARLRLPELHSAERYLTPEPEKELKLDKSSSSDLKTSPTIPSDTLS 899
QY 900 AETERHTSLGPPHPQVNFERSQCAIVLGKNSHFIGAVPLGSTEEDHESLGENVSPVE 959
Db 900 AETERHTSLGPPHPQVNFERSQCAIVLGKNSHFIGAVPLGSTEEDHESLGENVSPVE 959
QY 960 SDGIFEKERAHGASLTCKDVLKFNVISLVKTNKARVYLKTNRKIHIDDAALLTENRASA 1019
Db 960 SDGIFEKERAHGASLTCKDVLKFNVISLVKTNKARVYLKTNRKIHIDDAALLTENRASA 1019
QY 1020 TEMDKNTTASGLNHVSNWIKGPIGKNPLSSERGPSPELLTSSGSGKSVKGQSSGGRIRV 1079
Db 1020 TEMDKNTTASGLNHVSNWIKGPIGKNPLSSERGPSPELLTSSGSGKSVKGQSSGGRIRV 1079
QY 1080 AVEEELSKGEMMLPNSELTFLTNSADVQNDTHSQGKKSREMERREKLYQEKVDLP 1139
Db 1080 AVEEELSKGEMMLPNSELTFLTNSADVQNDTHSQGKKSREMERREKLYQEKVDLP 1139
QY 1140 VYTATGTKNFLRNIFHOSTEPVEGFDGSHAPVQDSRLNDSARAETHIAHSAIRE 1199
Db 1140 VYTATGTKNFLRNIFHOSTEPVEGFDGSHAPVQDSRLNDSARAETHIAHSAIRE 1199
QY 1200 EAPLEAPGNRTGPGPRSAVPRRVXQSLKQIRLPLEIKPERGVVLTATSTRWSESSPILQ 1259
Db 1200 EAPLEAPGNRTGPGPRSAVPRRVXQSLKQIRLPLEIKPERGVVLTATSTRWSESSPILQ 1259
QY 1260 GAKRNNLSLPLTLEMAGGQGISALGKSAAGPLASGKLEKAVLSSAGLSEASGKAFLP 1319
Db 1260 GAKRNNLSLPLTLEMAGGQGISALGKSAAGPLASGKLEKAVLSSAGLSEASGKAFLP 1319
QY 1320 KVRVHREDLLPQKTSNVSCAHDGQEIFLQKTRGPVNLKVNRPGRTPSKLGPMPKE 1379
Db 1320 KVRVHREDLLPQKTSNVSCAHDGQEIFLQKTRGPVNLKVNRPGRTPSKLGPMPKE 1379
QY 765 ----- 765

Db 1380 WESLEKSPKSTALRTKDIISLPLDRHESNHSIAKNEGQAEQTOREAAWTKOGGPRLCAP 1439
QY 765 -PPVLKRHQREITRTLQSDQEEIDYDDTISVEKKEDFDIYEDENQSPRSFOKTRHY 823
Db 1440 KPPVLRHQROISLPTFQPEEDKMDYDIDFSTETKGEDFDIYEDENQSPRSFOKTRHY 1499
QY 824 FIAAVERLMDYGMSSSPHYLRNRAQSGSVQFKVFOEFTDGSFTQPLYRGELNEHGL 883
Db 1500 FIAAVERLMDYGMSSSPHYLRNRAQSGSVQFKVFOEFTDGSFTQPLYRGELNEHGL 1559
QY 884 LCPYIRAEVEDNIMVTFERNQASRPYSFYSSLSIYEDQROGAEPKRNFKPNETKYFWK 943
Db 1560 LCPYIRAEVEDNIMVTFERNQASRPYSFYSSLSIYEDQROGAEPKRNFKPNETKYFWK 1619
QY 944 VOHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLLIGPLVCHTNTLNPAGROVTVQEFAL 1003
Db 1620 VOHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLLIGPLVCHTNTLNPAGROVTVQEFAL 1679
QY 1004 FFTIFDETKSWYFTENMERNCRAPCNIOEMEDPTKENYRFAHNGYIMDTLPLGLVMAQDO 1063
Db 1680 FFTIFDETKSWYFTENMERNCRAPCNIOEMEDPTKENYRFAHNGYIMDTLPLGLVMAQDO 1739
QY 1064 RIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMAVNLNYPGVFETVEMLPKAGIWRV 1123
Db 1740 RIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMAVNLNYPGVFETVEMLPKAGIWRV 1799
QY 1124 ECLIGEHLHAGNSTLFLVYSNCKQTPGLMASGHIRDFQITASGOYGOWAPKLARLHYSGS 1183
Db 1800 ECLIGEHLHAGNSTLFLVYSNCKQTPGLMASGHIRDFQITASGOYGOWAPKLARLHYSGS 1859
QY 1184 INAWSTKEPFWIKVDLAPMIHIGIKTOGAROKFSSLYISQFIIMYSLDGKKWQTYRGN 1243
Db 1860 INAWSTKEPFWIKVDLAPMIHIGIKTOGAROKFSSLYISQFIIMYSLDGKKWQTYRGN 1919
QY 1244 STGTLMVFFGNVDSGKHNFPPIIARYRHLPHTHYSIRSTLRLMELMGCCLNSCSMPL 1303
Db 1920 STGTLMVFFGNVDSGKHNFPPIIARYRHLPHTHYSIRSTLRLMELMGCCLNSCSMPL 1979
QY 1304 GMEKASIDAOITASSYFTNMATWSPSKARLHLOGRSNAMPQVNNPKENLOYDFOKTM 1363
Db 1980 GMEKASIDAOITASSYFTNMATWSPSKARLHLOGRSNAMPQVNNPKENLOYDFOKTM 2039
QY 1364 KYTGVTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKVKVFQGNQDSTFPVNSLD 1423
Db 2040 KYTGVTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKVKVFQGNQDSTFPVNSLD 2099
QY 1424 PPLTRYLRHPOSWVHQAIRMEVLGCEADLY 1457
Db 2100 PPLTRYLRHPOSWVHQAIRMEVLGCEADLY 2133

RESULT 3
FAB_MOUSE
ID FAB_MOUSE STANDARD; PRT; 2319 AA.
AC Q06194;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE COAGULATION FACTOR VIII PRECURSOR (PROCOAGULANT COMPONENT).
GN CF8 OR F8C.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIURONATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6 X CBA; TISSUE-LIVER;
RX MEDLINE; 93300511.
RA ELDER B., LAKICH D., GITSCHIER J.;
RT "Sequence of the murine factor VIII cDNA."
RL GENOMICS 16:374-379(1993).
CC -1- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS
CC AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE
CC ACTIVATED FORM, FACTOR XA.

CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: FOUND IN MOST TISSUES.
CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR V.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; L05573; G192457; -.
DR PIR; A47004; A47004.
DR MGD; MGI:88383; CF8.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PFAM; PF00394; Cu-oxidase; 3.
DR PFAM; PF00754; F5_F8_type_C; 2.
DR HSSP; P00451; 1CFG.
DR BLOOD COAGULATION; REPEAT; PLASMA; ACUTE PHASE; CALCIUM;
KW SIGNAL; GLYCOPROTEIN; SULFATATION.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2319 COAGULATION FACTOR VIII.
FT DOMAIN 20 349 F5/8 TYPE A 1.
FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.
FT DOMAIN 207 349 PLASTOCYANIN-LIKE 2.
FT DOMAIN 399 730 F5/8 TYPE A 2.
FT DOMAIN 399 573 PLASTOCYANIN-LIKE 3.
FT DOMAIN 583 730 PLASTOCYANIN-LIKE 4.
FT DOMAIN 760 1640 B.
FT DOMAIN 1683 2008 F5/8 TYPE A 3.
FT DOMAIN 1683 2008 PLASTOCYANIN-LIKE 5.
FT DOMAIN 1855 2008 PLASTOCYANIN-LIKE 6.
FT DOMAIN 2008 2156 F5/8 TYPE C 1.
FT DOMAIN 2161 2313 F5/8 TYPE C 2.
FT SITE 391 392 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 759 760 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1678 1679 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1324 1325 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
FT SITE 1641 1641 CLEAVAGE (ACTIVATION) (BY SIMILARITY).
FT MOD_RES 367 367 SULFATATION (BY SIMILARITY).
FT MOD_RES 737 737 SULFATATION (BY SIMILARITY).
FT MOD_RES 738 738 SULFATATION (BY SIMILARITY).
FT MOD_RES 742 742 SULFATATION (BY SIMILARITY).
FT MOD_RES 1669 1669 SULFATATION (REQUIRED FOR VWF BINDING)
FT MOD_RES 1669 1669 (BY SIMILARITY).
FT MOD_RES 1687 1687 SULFATATION (BY SIMILARITY).
FT DISULFID 173 199 PROBABLE.
FT DISULFID 547 573 PROBABLE.
FT DISULFID 1819 1845 PROBABLE.
FT DISULFID 2008 2156 BY SIMILARITY.
FT DISULFID 2161 2313 BY SIMILARITY.
FT CARBOHYD 61 61 POTENTIAL.
FT CARBOHYD 233 233 POTENTIAL.
FT CARBOHYD 259 259 POTENTIAL.
FT CARBOHYD 423 423 POTENTIAL.
FT CARBOHYD 601 601 POTENTIAL.
FT CARBOHYD 880 880 POTENTIAL.
FT CARBOHYD 958 958 POTENTIAL.
FT CARBOHYD 1015 1015 POTENTIAL.
FT CARBOHYD 1022 1022 POTENTIAL.
FT CARBOHYD 1026 1026 POTENTIAL.
FT CARBOHYD 1044 1044 POTENTIAL.
FT CARBOHYD 1076 1076 POTENTIAL.
FT CARBOHYD 1087 1087 POTENTIAL.
FT CARBOHYD 1136 1136 POTENTIAL.
FT CARBOHYD 1161 1161 POTENTIAL.
FT CARBOHYD 1192 1192 POTENTIAL.

FT CARBOHYD 1255 1255 POTENTIAL.
FT CARBOHYD 1268 1268 POTENTIAL.
FT CARBOHYD 1273 1273 POTENTIAL.
FT CARBOHYD 1274 1274 POTENTIAL.
FT CARBOHYD 1302 1302 POTENTIAL.
FT CARBOHYD 1316 1316 POTENTIAL.
FT CARBOHYD 1340 1340 POTENTIAL.
FT CARBOHYD 1378 1378 POTENTIAL.
FT CARBOHYD 1797 1797 POTENTIAL.
FT CARBOHYD 2105 2105 POTENTIAL.
SQ SEQUENCE 2319 AA; 266148 MW; E6014B4E CRC32;

Query Match 80.3%; Score 6261; DB 1; Length 2319;
Best Local Similarity 53.5%; Pred. No. 0;
Matches 1244; Conservative 97; Mismatches 109; Indels 876; Gaps 9;

QY 1 MQIELSTCFELCLLRFCSATRRYYIGAVELSWDYMQSD-IGELPYDARPPRPVPSFPF 59
DB 1 MQIALFACFFLSLENFNCSSAIRYYLIGAVELSWNYIQSDLSVLHTDSRFLPRMSTSPF 60
QY 60 NTSVYKKTLFEVEFTDHLFNIAKPRPVMGILPTIQAEVYDTVVITLKNMASHPVSLHA 119
DB 61 NTSIMYKKTVEYEYKDQLFNIAKPRPVMGILPTIWEVDTVVITLKNMASHPVSLHA 120
QY 120 VGVSYWKASEGAEYDDQTSQREKEDDKVPFGSHYVWQVLKENGPMASDPLCTYSYLS 179
DB 121 VGVSYWKASEGEYEDQTSQMEKEDDKVPFGSHYVWQVLKENGPMASDPLCTYSYMS 180
QY 180 HYDLVKDLNSGLIGALLVCRGSLAKEKTQTLHKFILLFAVEDEKSWHSETKNSLMQDR 239
DB 181 HYDLVKDLNSGLIGALLVCRGSLAKEKTQTLHKFILLFAVEDEKSWHSETKNSLMQDR 240
QY 240 DAASARAWPKMHTVNGVYVNRSLPGLIGCHRSKYVWHVIGMGTPEVHSIFLEGHTFLVRN 299
DB 241 DSASARDWPKMHTVNGVYVNRSLPGLIGCHRSKYVWHVIGMGTPEVHSIFLEGHTFLVRN 300
QY 300 HROASLEISPIFTFLTAQTLMDLGOELLFCHISSHOHDMGEAYVKVDSCEPEQOLRMK-N 358
DB 301 HROASLEISPIFTFLTAQTLMDLGOELLFCHISSHOHDMGEAYVKVDSCEPEQOLRMK 360
QY 359 NEEAEDYDDDLTDESEMDVVRFPDDNSPSFIQIRSVAKKHPKTVWHYIAAEEDWDYAPLV 418
DB 361 NEEAEDYDDDLT-SEMDMTLDYDSSP-FIQIRSVAKKHPKTVWHYIAAEEDWDYAPSV 418
QY 419 LAPDRSYKSQYLNNGPQIRGRKKYKVRFMAYVDEFTKTRERAIQHESGILGPLLYGEVGD 478
DB 419 PTDNGSYKSQYLNNGPQIRGRKKYKVRFMAYVDEFTKTRERAIQHESGILGPLLYGEVGD 478
QY 479 TLLIFKNQASRPYNIYPHGITVDRPLYSRRLPKGVKHLKDFPLPGEIFKYKTVTVED 538
DB 479 TLLIFKNQASRPYNIYPHGITVDRPLYSRRLPKGVKHLKDFPLPGEIFKYKTVTVED 538
QY 539 GPTKSDPRCLTRYSSSFVNMERDLASGLIGPLLCYKESVDQRGNOIMSDKRNVLFSVF 598
DB 539 GPTKSDPRCLTRYSSSFVNMERDLASGLIGPLLCYKESVDQRGNOIMSDKRNVLFSVF 598
QY 599 DENRSWYLTENIQRLPNPAGVQLEDEPFOASNIMHSINGVYFDSLQSVCLHEVAYWYI 658
DB 599 DENRSWYLTENIQRLPNPAGVQLEDEPFOASNIMHSINGVYFDSLQSVCLHEVAYWYI 658
QY 659 LSIQAQTDLSVFSFGYTFKHKMYVEDTLTLPFSGETVFMSEMPGLMWLGCHNSDFRN 718
DB 659 LSIQAQTDLSVFSFGYTFKHKMYVEDTLTLPFSGETVFMSEMPGLMWLGCHNSDFRN 718
QY 719 RGMFALLKVSCKDKTGDYEDYSAYLLSKNNAIEPRFSQN----- 765
DB 719 RGMFALLKVSCKDKTGDYEDYSAYLLSKNNAIEPRFSQN----- 765
QY 765 ----- 765
DB 779 TIPKNDMEKIEPQFEIEAEMLKVQSVSVSDMLMLGQSHPTPHGLFLSDGQAEIYEAIHD 838

QY 765 ----- 765
Db 839 DHSPNAIDSEGPSKVTLQRPRESHSEKIVFTPPGLQLRSNKSLETTIEVKWKGLQV 898
QY 765 ----- 765
Db 899 SSLPSNLMTTTLSDNLKATFEKTDSSGFPDMPVHSSSKLSTAFGKKAYSLVGSHPVLN 958
QY 765 ----- 765
Db 959 ASEENSDSNILDSTLMYSQESLPRDNLISJENDRLREKRFGIALTLTKDNTLFKDNVSL 1018
QY 765 ----- 765
Db 1019 MKNKYNHSTNEKLHTESPSTIENSTDLQDAILKVNSEIQEVTALIHDTLIGKNST 1078
QY 765 ----- 765
Db 1079 YLRNLHMLNRTSTKNKDIHRKDEDPIDQDEENTIMPFKMLFLSESSNWEKKTNGNNS 1138
QY 765 ----- 765
Db 1139 LNSEQHSPKQLVYLMFKYVKQNSFLSEKNKVTEQDGFKNIGLKDMAFPNMSIFLT 1198
QY 765 ----- 765
Db 1199 TLSNVHENGRHNOEKNIOEIEKEALIEKVVLPQVHEATGSKNFLKDILILGTRQNISL 1258
QY 765 ----- 765
Db 1259 YEYHVPLQNTSINNSTNTVOIHMEHFRRKDKETNSEGLVNTKREYKNYPSOKNIT 1318
QY 765 ----- 765
Db 1319 TQSKRALGQFRLSTOWLKTINCSTOCIIKQIDHSKEMKFKITKSLSDSSVIKSTTQTN 1378
QY 765 ----- 765
Db 1379 SSDSHIVKTSAPFPIDLKSPFQNKFSHVQASSIYDFKTKSSRIOESNMLKETKINNP 1438
QY 765 ----- 765
Db 1439 SLALPWNMFIDQKFTSPGKSNTNSVYKKRENIIIFLKPTLPESGKIELLPQVSIQEE 1498
QY 765 ----- 765
Db 1499 EILPTETSHGSPGHLNLMKEVFLQKIQGPTKWNKAKRHGESIKGTESKNTRSKLLNH 1558
QY 765 ----- 765
Db 1559 AMDYHYAAQIPKDMWKSKEKSPEIISIKQEDTILSLRPHGNSHIGANEKONWPQRETW 1618
QY 765 ----- 765
Db 1619 VKOGQTRTCSQIPVLRKHQREL--SAFQSEQEATDYDAITIE-TIEFEDITYSEDIKQ 1675
QY 812 SPRSFQKTRHYFIAAVERLMDYGMSSSPHYLRNRAQSGSVQFKVVFQEFDTGSGFTOP 871
Db 1676 GPRSFQKTRHYFIAAVERLMDYGMSTS-HYLRNRQSDNVQFKKVFQEFDTGSGFSQP 1734
QY 872 LYRGELNEHGLLGPYIRAEVEDNIMVTFRQASRPYSFYSSLSISYEEDROGAEPKRN 931
Db 1735 LYRGELNEHGLLGPYIRAEVEDNIMVTFKQASRPYSFYSSLSISYKEDOR-GEPPRRNF 1793
QY 932 VKPNETKITFWKVQHHMARTKDEFDCKAMAYFSDVDLEKDVHSGLIGPLLVCHTNTLNP 991
Db 1794 VKPNETKITFWKVQHHMARTKDEFDCKAMAYFSDVDLERDMSGLIGPLLVCHTNTLNP 1853
QY 992 HGRQVTVQEFALFTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFFHAINGYIM 1051
Db 1854 HGRQVTVQEFALFTIFDETKSWYFTENYKRNCKTPCNFQMEDPTLKENYRFFHAINGYIM 1913
QY 1052 DTLPLVMAQDQIRIRWYLLSMGNSNENIHSIHSGHVFTVRKKEEYKMALYNLYPGVFETV 1111

Db 1914 DTLPLVMAQDQIRIRWYLLSMGNNENIOSIHFSGHVFTVRKKEEYKMAVYNLYPGVFETL 1973
QY 1112 EMLPSKAGIWRVECLIGELHAGMSTLELVYSNKCQTPPLGMASGHIRFQITASGOYGM 1171
Db 1974 EMIPSRAGIWRVECLIGELHAGMSTLELVYSKOCQPLGMASGSIRFQITASGHYGM 2033
QY 1172 APKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGAROKESSLYISQFIIMS 1231
Db 2034 APNLARLHYSGSINAWSTKEPFSWIKVDLLAPMIVHGIKTQGAROKESSLYISQFIIMS 2093
QY 1232 LDGKKWQTYRGNSTGTLMVFFGNVDSGSIKHNIFNPPIARIYIRLHPHYISIRSTLMEL 1291
Db 2094 LDGKKWQTYRGNSTGTLMVFFGNVDSGSIKHNIFNPPIARIYIRLHPHYISIRSTLMEL 2153
QY 1292 MGC DLNSGSMPLGMEKSAISDAQITASSYFTMTFATWSPSKARLHLQGRSNAMRPQVNDP 1351
Db 2154 MGC DLNSGSMPLGMEKSAISDAQITASSYFTMTFATWSPSKARLHLQGRSNAMRPQVNDP 2213
QY 1352 KEWLQVDFQKTMKVGTGTQGVKSLTSMYVKEFLISSQDGHQWTFEQNGKVKVFOGN 1411
Db 2214 KQWLQVDFQKTMKVGTGTQGVKSLTSMYVKEFLISSQDGHQWTFEQNGKVKVFOGN 2273
QY 1412 QDSFTPVVNSLDPPLTRYLRIHPQSVWQIALRMEVLCGEAODLY 1457
Db 2274 QDSFTPVVNSLDPPLTRYLRIHPQSVWQIALRMEVLCGEAODLY 2319

RESULT 4
FA5_HUMAN
ID FA5_HUMAN STANDARD; PRT: 2224 AA.
AC P12259; Q14285;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE COAGULATION FACTOR V PRECURSOR (ACTIVATED PROTEIN C COFACTOR).
GN F5.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92232668.
RA CRIPE L.D., MOORE K.D., KANE W.H.;
RT "Structure of the gene for human coagulation factor V.";
RL BIOCHEMISTRY 31:3777-3785(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87260886.
RA JENNY R.J., PITTMAN D.D., TOOLE J.J., KRIZ R.W., ALDABE R.A.,
RT HEWICK R.M., KAUFMAN R.J., MANN K.G.;
RL "Complete cDNA and derived amino acid sequence of human factor V.";
RL PROC. NATL. ACAD. SCI. U.S.A. 84:4846-4850(1987).
RN [3]
RP SEQUENCE OF 1-1600 FROM N.A.
RX MEDLINE; 88107560.
RA KANE W.H., ICHINOSE A., HAGEN F.S., DAVIE E.W.;
RT "Cloning of cDNAs coding for the heavy chain region and connecting
RT region of human factor V, a blood coagulation factor with four types
RT of internal repeats.";
RL BIOCHEMISTRY 26:6508-6514(1987).
RN [4]
RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
RX MEDLINE; 86313665.
RA KANE W.H., DAVIE E.W.;
RT "Cloning of a cDNA coding for human factor V, a blood coagulation
RT factor homologous to factor VIII and ceruloplasmin.";
RL PROC. NATL. ACAD. SCI. U.S.A. 83:6800-6804(1986).
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RC TISSUE-FIBROBLAST;
RX MEDLINE; 93203619.
RA SHEN N.L.L., FAN S.-T., PYATI J., GRAFF R., LAPOLLA R.J.,

RA EDGINGTON T.S.;
RT "The serine protease cofactor factor V is synthesized by
RT lymphocytes."
RL J. IMMUNOL. 150:2992-3001(1993).
RN [6]
RP VARIANT APCR GLN-534.
RX MEDLINE; 94217810.
RA BERTINA R.M., KOELEMAN B.P.C., KOSTER T., ROSENDAAL F.R.,
RA DIRVEN R.J., DE RONDE H., VAN DER VELDEN P.A., REITSMA P.H.;
RT "Mutation in blood coagulation factor v associated with resistance to
RT activated protein C."
RL NATURE 369:64-67(1994).
CC -I- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
CC -I- SUBUNIT: FACTOR VA IS COMPOSED OF AN HEAVY CHAIN AND OF A LIGHT
CC CHAIN NONCOVALENTLY BOUND. THE INTERACTION BETWEEN THE TWO CHAINS
CC IS CALCIUM-DEPENDENT.
CC -I- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
CC REPEATS.
CC -I- PTM: THROMBIN ACTIVATES FACTOR V PROTEOLYTICALLY TO THE ACTIVE
CC COFACTOR, FACTOR V(A) (FORMATION OF A HEAVY CHAIN AT THE N-
CC TERMINUS AND A LIGHT CHAIN AT THE C-TERMINUS).
CC -I- DISEASE: OMREN PARAHEMOPHILIA, AN HEMORRHAGIC DIATHESIS, IS DUE
CC TO A DEFICIENCY OF FACTOR V. OTHER DEFECTS IN F5 RESULTS IN A
CC FORM OF THROMBOPHILIA KNOWN AS APC RESISTANCE (APCR). THE APCR
CC MUTATION IS FOUND IN ABOUT 5% OF THE POPULATION WHICH SUGGEST THAT
CC A SLIGHT THROMBOTIC TENDENCY MAY CONFER SOME ADVANTAGE IN FETAL
CC IMPLANTATION.
CC -I- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -I- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -I- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L32779; G488110; -
DR EMBL; L32755; G488110; JOINED.
DR EMBL; L32756; G488110; JOINED.
DR EMBL; L32757; G488110; JOINED.
DR EMBL; L32758; G488110; JOINED.
DR EMBL; L32759; G488110; JOINED.
DR EMBL; L32760; G488110; JOINED.
DR EMBL; L32761; G488110; JOINED.
DR EMBL; L32762; G488110; JOINED.
DR EMBL; L32763; G488110; JOINED.
DR EMBL; L32764; G488110; JOINED.
DR EMBL; L32765; G488110; JOINED.
DR EMBL; L32766; G488110; JOINED.
DR EMBL; L32767; G488110; JOINED.
DR EMBL; L32768; G488110; JOINED.
DR EMBL; L32772; G488110; JOINED.
DR EMBL; L32773; G488110; JOINED.
DR EMBL; L32774; G488110; JOINED.
DR EMBL; L32775; G488110; JOINED.
DR EMBL; L32776; G488110; JOINED.
DR EMBL; L32777; G488110; JOINED.
DR EMBL; L32778; G488110; JOINED.
DR EMBL; M16967; G182412; -
DR EMBL; M14335; G182798; -
DR PIR; A25897; A25897.
DR PIR; A28028; A28028.
DR MIM; 134400; -
DR MIM; 188055; -
DR MIM; 227310; -

DR MIM; 227400; -
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PFAM; PF00394; Cu-oxidase; 3.
DR PFAM; PF00754; F5_F8_type_C; 2.
DR HSSP; P00450; 1KCM.
KW BLOOD COAGULATION; PLASMA; GLYCOPROTEIN; CALCIUM; SIGNAL; ZYMOGEN;
KW REPEAT; POLYMORPHISM; DISEASE MUTATION; THROMBOPHILIA.
FT SIGNAL 1 28
FT CHAIN 29 2224
FT CHAIN 29 737
FT PEPTIDE 738 1573
FT CHAIN 1574 2224
FT DOMAIN 30 329
FT DOMAIN 30 193
FT DOMAIN 203 329
FT DOMAIN 348 684
FT DOMAIN 348 526
FT DOMAIN 536 684
FT DOMAIN 692 1573
FT DOMAIN 895 928
FT DOMAIN 895 911
FT REPEAT 912 928
FT REPEAT 1135 1148
FT DOMAIN 1185 1463
FT REPEAT 1185 1193
FT REPEAT 1194 1202
FT REPEAT 1203 1211
FT REPEAT 1212 1220
FT REPEAT 1221 1229
FT REPEAT 1230 1238
FT REPEAT 1239 1247
FT REPEAT 1248 1256
FT REPEAT 1257 1265
FT REPEAT 1266 1274
FT REPEAT 1275 1283
FT REPEAT 1284 1292
FT REPEAT 1293 1301
FT REPEAT 1302 1310
FT REPEAT 1311 1319
FT REPEAT 1320 1328
FT REPEAT 1329 1337
FT REPEAT 1338 1346
FT REPEAT 1347 1355
FT REPEAT 1356 1364
FT REPEAT 1365 1373
FT REPEAT 1374 1382
FT REPEAT 1383 1391
FT REPEAT 1392 1400
FT REPEAT 1401 1409
FT REPEAT 1410 1418
FT REPEAT 1419 1427
FT REPEAT 1428 1436
FT REPEAT 1437 1445
FT REPEAT 1446 1454
FT REPEAT 1455 1463
FT REPEAT 1464 1472
FT REPEAT 1473 1481
FT REPEAT 1482 1490
FT REPEAT 1493 1501
FT DOMAIN 1578 1907
FT DOMAIN 1578 1751
FT DOMAIN 1761 1907
FT DOMAIN 1907 2061
FT DOMAIN 2066 2221
FT SITE 737 738
FT SITE 1046 1047
FT SITE 1573 1574
FT DISULFID 167 193
FT DISULFID 500 526
FT DISULFID 1725 1751
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.
13.
14.
15.
16.
17.
18.
19.
20.
21.
22.
23.
24.
25.
26.
27.
28.
29.
30.
31.
32.
33.
34.
35.
F5/8 TYPE A 3.
PLASTOCYANIN-LIKE 5.
PLASTOCYANIN-LIKE 6.
F5/8 TYPE C 1.
F5/8 TYPE C 2.
CLEAVAGE (BY THROMBIN).
CLEAVAGE (BY THROMBIN).
PROBABLE.
PROBABLE.
PROBABLE.

FT DISULFID 1907 2061 BY SIMILARITY.
FT DISULFID 2066 2221 BY SIMILARITY.
FT CARBOHYD 51 51 POTENTIAL.
FT CARBOHYD 55 55 POTENTIAL.
FT CARBOHYD 239 239 POTENTIAL.
FT CARBOHYD 297 297 POTENTIAL.
FT CARBOHYD 382 382 POTENTIAL.
FT CARBOHYD 460 460 POTENTIAL.
FT CARBOHYD 468 468 POTENTIAL.
FT CARBOHYD 554 554 POTENTIAL.
FT CARBOHYD 741 741 POTENTIAL.

Query Match 30.7%; Score 2390.5; DB 1; Length 2224;
Best Local Similarity 26.1%; Pred. NO. 2.6e-146;
Matches 593; Conservative 276; Mismatches 484; Indels 923; Gaps 35;

QY 22 RRYVIGAVELSWDMQSDLGELPYDARFPPRVPRKSPFNTSVV-YKKTLEVEFTDHLFNI 80
DB 32 RQFYVAAQGISWSYRPE-----PNNSSLNLSTVTSFKKIYREYEBY-FKK 75
QY 81 AKRPRPWWGLGPTIQAEVYDTVVITLKNMASHPVSLHAGVSYWKASEGAEYDDQTSOR 140
DB 76 EKQOSTISGLGPTLYAEVGDIIKHFKNKADKPLSIHQGIRYSKLSGASYLDHTFPA 135
QY 141 EKEDDKVFPGSGSHYVMOVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCRE 200
DB 136 EKMDDAVAPGREYTYEWSISEDSEDTDDPCLFTHIYSHENLIEDFNGLIGPLICK 195
QY 201 GSLAKEKTO-TLHK-FILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVN 258
DB 196 GTLEGGTQKTFKQIVLLFAVDESKSMSQSS-----LMYTVNGYVN 239
QY 259 RSLPGLIGCHRSYVWHVIGMGTPEVHSIFLEGHTFLVRNHRQASLEISPTFLAQT 318
DB 240 GTMDITVCAHDHISWHLIGMSSGDELFSIHNGQVLEQNHKYSAILTVSATSTTANMT 299
QY 319 LMDIGQFLLFCHISSHQDGMAYKVDSCPEEPOLRMKNNEAEYDDDLTDSMDVVR 378
DB 300 VGPEGKWIISLTFKHLQAGMOAYIDIKNCPRKTRNLKKTRE----- 343
QY 379 FDDDNPSFIOIRSVAKKHPTWVHYIAAEEDMDYAPLYLAPDDRYSKQYLNGPQRI 438
DB 343 -----QRHMKRWEYFIAAEVIMDYAPVIPANMDKKYRSQHLNFSNOI 387
QY 439 GRKKKKVRFMAYTDEFTKREAT---QHESGILGPLLYGEVGDITLLIFKNQASRPYNIY 495
DB 388 GKHKKKVMYQYEDEF-TKHTVNMKEDGILGPIIRAQVRDTLKIYFKNMASRPYSIY 446
QY 496 PHGTT-----DYRPLYSRRLPKGYKHLKDFILPGEIFKYKWTYVEDGPTKSPRCLT 549
DB 447 PHGVTFSPYEDEVNSSFT---SGNNMTIRAVDPGETTYTKWNILEFDEPTENDACCLT 502
QY 550 RYSSSVNMERDLASGLIGPLLYCKESVDQRGNOIMSDKRNVLFSVEDENRSWYLTEN 609
DB 503 RPYSDVDIMRDIASGLIGLLICKRSRLDRGIDRAADIEQQAFAVDEKMSWYLEDN 562
QY 610 IQEFLPNPAGVQLEDEPEQASNMHSINGYVFDL-QLSVCLHEVAYWYILSIGAQTDFL 668
DB 563 INKFCENPDEVKRDPKFYESNIMSTINGYVPESTITLGFCEFDITVQWHFCSVGTONEIL 622
QY 669 SVFSSGYTFKHMYEDTLTFPESGETVFMSENGMLILGCHNS----- 715
DB 623 TIHFTGHSFIYGRKHEDTLTLFPMGESVTVTMDVGTWMLTSMSSPRSKRLRFRDY 682
QY 715 -----DFRNR----- 720
DB 683 KCIIPDDEDSYEIEPPESTVMATRKMHDRLEPEDEESDADYDQNRLLAALGIRSFRRNS 742
QY 720 -----GMTAL----- 725
DB 743 SLNQEEEFNLALALENGTEFVSSNTDIIVGSNTYSPSNISKFTVNNLAEPQKADSHQ 802
QY 725 -----LKVSSCDKNTGDIYEDSYED----- 745

DB 803 ATTAGSPLRHLIGKNSVLSNSTAEHSSPYSEDPIEDPLOPDVYTGIRLLSLGAGEFKSOEH 862
QY 745 ----- 745
DB 863 AKHGPVERDQAAKHFRFSWKKLLAHKVGRHLSQDTGSPSGMRPWEDLPsODTGSPPSRMR 922
QY 745 -----ISAYLSKNN----- 756
DB 923 PKKDPSSDLLLKQSNSSKILVGRWHLASEKSYEIIQDTDEDTAVNNMILISQONASRAW 982
QY 756 ----- 756
DB 983 GESTPLANKPGKQSGHPKFPYVRHKSLOVRQDQCGKSRLLKKSQFLIKTRKKKEKHTHAP 1042
QY 756 IEPRSF-----SQ 763
DB 1043 LSPRTFHPLRSEAYNTFSERLKLHSLVLAHKSNETSLPTDLNQTLPSMDFGWLASLPDHNO 1102
QY 764 N----- 765
DB 1103 NSSNDTGQASCPPGLYQTVPEEHYQTFPIODPDQMHSTSDPSHRSSSPLESEMLEYDRS 1162
QY 765 -----PPVLRH- 772
DB 1163 HKSFPPTDISQSPSSSEHEWQTVISPDLSQVTLSPELSQTNLSPDLSHHTLSPELIQRL 1222
QY 772 ----- 772
DB 1223 SPALQMPISPDLSHTTLLSPDLSHHTLLSLDLSQTNLSPELSQTNLSPALQMPISPDLSH 1282
QY 772 ----- 772
DB 1283 TLLSLDFSQTNLSPELSHMTLSPELSQTNLSPALQMPISPDLSHTTLLSLDFSQTNLSPE 1342
QY 772 -----QREITRTTLQSDQEEI----- 788
DB 1343 LSQTNLSPALQMPISPDPSHTTLLSLDLSQTNLSPELSQTNLSPDLSEMPPLADLSQIPL 1402
QY 788 ----- 788
DB 1403 TPDLQMTLSPDLGETDLSPNFGQMSLSPDLSQVTLSPDISDTLLPDLSQLSPDDLQ 1462
QY 788 -----DYDDT 792
DB 1463 IFYPSSESSQSLLEFNESFPYPDLGQMPSPSPPTLNDTFLSKENPLVIYVGLSKDGTDY 1522
QY 793 ISYEMKE-----DFDIYDE-----DENQSP-----RSFOKTRHY 823
DB 1523 IEIIPKEEVQSSSEDYAIEDYVYDDPYKTDVRTNINSSRDPDNIAMWYLRNNGNRNRY 1582
QY 824 FIAAVERLMDYGMSSSPHYLRNR--AQSGSVPQ--FKKVQOEFTDGSFTQPLYRGELN 878
DB 1583 YIAAEISWDY-----SEFVQRETDIEDSDDIPEDTTYKKVFRKYLDDSTFTKRDPRGEYE 1638
QY 879 EHLGLGPYIRAEVDNIMVTRNQASRPYSYSSLSYE-----EDQGAEPKRN 930
DB 1639 EHLGILGPYIRAEVDVIOVFRKFLASRPYSLAHAGLSYEKSSGKTYEEDDSEWFKEDN 1698
QY 931 FVAPNETKTYFMVQHHMAPTKDEFDCAKAWAFSDVDLEKDVHSGLLIGPLVCHTNTLNP 990
DB 1699 AVQPNSSYTYVWHAITERSGPESGSACRAMAYSAVNPEKDIHSGLLIGPLLCOKGILHK 1758
QY 991 AHGRQVTVQEPALFETIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRHAINGYI 1050
DB 1759 DSNMPVDMREFVLLFMTFDEKKSWMYEEKSRSSWR-----LTSSEMKKSHEHAINGMI 1812
QY 1051 MDLPLGLVMAQDQIRRWYLLSNGSNENIHSIHSGHVFTVRKKEEYKMALYNLYPGVFET 1110
DB 1813 Y-SLPGLKMYEQEWVRLHLNLNIGSODIHVYHFGQTLLENGNKQHLGVPPLPGSFKT 1871
QY 1111 VEMLPKAGIWRVECLLIGEHLHAGMSTLFLVYSNKCQTPLGMA SGHIRDFQTASGOYGO 1170

```
DB 1872 LEMKASKPGWMLNTEVEGENORAGMOTPLIMDRDCRMPMGLSTGIISSQIKASEFLGY 1931
QY 1171 WAEKLARLHYSGSINAWSTKE--PES--WIKYDLLAPMIHIGIKTOGAROFSSLYIS 1224
DB 1932 WEPLRLANNGSYNAWSVEKLAFAASKPWIOYDMQKEVITGTQOGAKHYKSCYTT 1991
QY 1225 QETIMYSLDGKKMOTYRGNSTGTLMVEFGNVDSGKIKHNFNPPIIARYIRLPHYSIR 1284
DB 1992 EFVYAYSSNOINWQIFKGNSTRNMYENGNSDASTIKENQFDPIVARYIRISPTRAYNR 2051
QY 1285 STLRLMELMGCDLNSCMLPGMESKASIDAQITASSYFTNMFAT-WSPSKARLHQGRSNA 1343
DB 2052 PTLRLLEOGCEVNGSTPLGMEKNGIKENKQITASSPKKSWMGDYWEPRARLNAQGRVNA 2111
QY 1344 WRPOVNNPKEWLQVDFOKTMKVTGVTQGVKSLTSMYVKEFLISSQDGHOWTLEFQNG 1403
DB 2112 WQAKANNKKOWLEIDLKIKKITAITQGCKSLSEMYKSYTIHYSEQGVEMKPYRLKS 2171
DB 1404 KV--KVFQGNODSFTPVNSLDPPLITRYLRIPHQSWVYQIALRMEVLGCEAODLY 1457
DB 2172 SMVDKIFEGNTNTKGHVKNFNPPIISRFIRVIPKFWNQSITLRLLEFGC--DIY 2224

RESULT 5
FA5_BOVIN STANDARD; PRT: 2211 AA.
AC Q28107; Q28108;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE COAGULATION FACTOR V PRECURSOR (ACTIVATED PROTEIN C COFACTOR).
GN F5.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDEA; BOVIDAE; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 92147638.
RA GUINTO E.R., ESMON C.T., MANN K.G., MAGILLIYRAY R.T.;
RT "The complete cDNA sequence of bovine coagulation factor V.";
RL J. BIOL. CHEM. 267:2971-2978(1992).
CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
CC -1- SUBUNIT: FACTOR VA IS COMPOSED OF AN HEAVY CHAIN AND OF A LIGHT
CC CHAIN NONCOVALENTLY BOUND. THE INTERACTION BETWEEN THE TWO CHAINS
CC IS CALCIUM-DEPENDENT.
CC -1- DOMAIN: DOMAIN B CONTAINS 29.5 X 9 AA TANDEM REPEATS, AND 2 X 17
CC AA REPEATS.
CC -1- PTM: THROMBIN ACTIVATES FACTOR V PROTEOLYTICALLY TO THE ACTIVE
CC COFACTOR, FACTOR V(A) (FORMATION OF A HEAVY CHAIN AT THE N-
CC TERMINUS AND A LIGHT CHAIN AT THE C-TERMINUS).
CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M81440; G163038; -
CC EMBL; M81441; G163040; -
CC PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
CC PFAM; PF00394; Cu-oxidase; 3.
CC PFAM; PF00754; F5_F8_type_C; 2.
CC HSSP; P00450; 1KCW.

KW BLOOD COAGULATION; PLASMA; GLYCOPROTEIN; CALCIUM; SIGNAL; ZMOGEN;
KW REPEAT.
FT SIGNAL 1 28
FT CHAIN 29 2211
FT CHAIN 29 741
FT PEPTIDE 742 1564
FT CHAIN 1565 2211
FT DOMAIN 30 327
FT DOMAIN 30 193
FT DOMAIN 203 327
FT DOMAIN 348 686
FT DOMAIN 348 525
FT DOMAIN 535 686
FT DOMAIN 696 1564
FT SIMILAR 899 915
FT DOMAIN 1124 1151
FT REPEAT 1124 1137
FT REPEAT 1138 1151
FT DOMAIN 1188 1453
FT REPEAT 1188 1196
FT REPEAT 1197 1205
FT REPEAT 1206 1214
FT REPEAT 1215 1223
FT REPEAT 1224 1232
FT REPEAT 1233 1241
FT REPEAT 1242 1250
FT REPEAT 1251 1259
FT REPEAT 1260 1268
FT REPEAT 1269 1277
FT REPEAT 1278 1286
FT REPEAT 1287 1295
FT REPEAT 1296 1304
FT REPEAT 1305 1313
FT REPEAT 1314 1322
FT REPEAT 1323 1331
FT REPEAT 1332 1340
FT REPEAT 1341 1349
FT REPEAT 1350 1358
FT REPEAT 1359 1367
FT REPEAT 1368 1376
FT REPEAT 1377 1385
FT REPEAT 1386 1394
FT REPEAT 1395 1403
FT REPEAT 1404 1412
FT REPEAT 1413 1421
FT REPEAT 1422 1430
FT REPEAT 1431 1439
FT REPEAT 1440 1444
FT REPEAT 1445 1453
FT DOMAIN 1569 1890
FT DOMAIN 1569 1738
FT DOMAIN 1748 1890
FT DOMAIN 1894 2048
FT DOMAIN 2053 2208
FT SITE 741 742
FT SITE 1034 1035
FT SITE 1564 1565
FT DISULFID 167 193
FT DISULFID 499 525
FT DISULFID 1712 1738
FT DISULFID 1894 2048
FT DISULFID 2053 2208
FT CARBOHYD 225 225
FT CARBOHYD 239 239
FT CARBOHYD 297 297
FT CARBOHYD 382 382
FT CARBOHYD 460 460
FT CARBOHYD 553 553
FT CARBOHYD 587 587
FT CARBOHYD 745 745
FT CARBOHYD 756 756

POTENTIAL.
COAGULATION FACTOR V.
HEAVY CHAIN (BY SIMILARITY).
ACTIVATION PEPTIDE (CONNECTING REGION)
(BY SIMILARITY).
LIGHT CHAIN (BY SIMILARITY).
F5/8 TYPE A 1.
PLASTOCYANIN-LIKE 1.
PLASTOCYANIN-LIKE 2.
F5/8 TYPE A 2.
PLASTOCYANIN-LIKE 3.
PLASTOCYANIN-LIKE 4.
B.
TO 17 AA REPEATS IN HUMAN FA5.
2 X 14 AA TANDEM REPEATS.
1.
2.
30 X 9 AA TANDEM REPEATS OF [AS]-L-S-P-
D-[LP]-[GS]-Q-[TE] (APPROXIMATE).
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.
13.
14.
15.
16.
17.
18.
19.
20.
21.
22.
23.
24.
25.
26.
27.
28.
29 (PARTIAL).
30.
F5/8 TYPE A 3.
PLASTOCYANIN-LIKE 5.
PLASTOCYANIN-LIKE 6.
F5/8 TYPE C 1.
F5/8 TYPE C 2.
CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
PROBABLE.
PROBABLE.
PROBABLE.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.
POTENTIAL.
```


FT CARBOHYD 774 774 POTENTIAL.
FT CARBOHYD 780 780 POTENTIAL.
FT CARBOHYD 902 902 POTENTIAL.
FT CARBOHYD 952 952 POTENTIAL.
FT CARBOHYD 964 964 POTENTIAL.
FT CARBOHYD 1044 1044 POTENTIAL.
FT CARBOHYD 1053 1053 POTENTIAL.
FT CARBOHYD 1062 1062 POTENTIAL.
FT CARBOHYD 1071 1071 POTENTIAL.
FT CARBOHYD 1078 1078 POTENTIAL.
FT CARBOHYD 1094 1094 POTENTIAL.
FT CARBOHYD 1451 1451 POTENTIAL.
FT CARBOHYD 1490 1490 POTENTIAL.
FT CARBOHYD 1550 1550 POTENTIAL.
FT CARBOHYD 1690 1690 POTENTIAL.
FT CARBOHYD 1839 1839 POTENTIAL.
FT CARBOHYD 1997 1997 POTENTIAL.
FT CARBOHYD 2196 2196 POTENTIAL.
FT VARIANT 587 592 NETLPA -> T (IN VARIANT 2).
SQ SEQUENCE 2211 AA; 248981 MW; 9B017C5C CRC32;

Query Match 30.4%; Score 2370; DB 1; Length 2211;
Best Local Similarity 26.2%; Pred. No. 5.4e-145;
Matches 594; Conservative 265; Mismatches 487; Indels 924; Gaps 35;

QY 22 RRYLGAVELSDMYQSDLGELPVDARPPRPVPSFPFNTSVYKKTLFVEETDHLFNTA 81
DB 32 ROFYVAQSIRWNYR-----PESTHL-----SSKPFETS--FKKIYREY-EAYFOKE 76
QY 82 KPRPWWGLGPTIOAEVYDTVTITLKNASHPVSLHAGVSYWKASEGAEYDQTSORE 141
DB 77 KQSRISGLGPTLYAEVGDIMKVFKNKHKPLSIHAQIKYSKSEGSYSDHTLPM 136
QY 142 KEDKVPFGSGSHTYVQVLENGPMASDPLCLTYSYLSHYDLVKDLNSGLIGALLVCREG 201
DB 137 KMDAVAPGQETTYEWIISHSGPTHDPCLTHIYSYVNLVEDFNSGLIGPLICKRG 196
QY 202 SLAKEKQTL--HKFLLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVR 259
DB 197 TLTEDGTQKMEKQHVLMFAVEDESKSWNQTS-----LMYTVNGYVNG 240
QY 260 SLPLGLGCHRRKSVYWHVIGNGTTPREVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTL 319
DB 241 TMPDITVCANDHISWHLIGSSGPELFSIHNGQVLEQNHKISAITLVSATSTANMTV 300
QY 320 MDLGQFLFCHISSHQHDCMEAYVKVDSCEEPQLRMKNNEEAEDYDDDLTDSMDVVR 379
DB 301 SPEGRWITASLIPRHFOAGNOAYIDIKNCAKTRNPKK-----LTRDQ----- 344
QY 380 DDDNSPSFIQIRSVAKKHKTWVHYIAAEEDWDYAPLVLPDDRYSKQYLNNGPORIG 439
DB 344 -----RRHKKWEYFIAAEVIMDYAPIIPANMDKYYRSLHLDNFSNRIG 388
QY 440 RYKRVRFMAVTDTEFKTR--EAIQHESGIGPLLYGEVGDTLIIIFKNQASRPYNTYPH 497
DB 389 KHYKKVYKQYQDDSFTRKLEDPSSSEGDIIPIRAQVDTLKIYFKNMSRSYSTYPH 448
QY 498 GIT-----DVRPLYSRLLPKGVKHLKDFPILPGEIFRYKWTVTVEGPTKSDPRCL 548
DB 449 GVTFSPIYDNEVNSSSTSGSNTMIRAVR-----PGETYTYKWNILESDEPTENDAQCL 500
QY 549 TRYSSFVNMRDLASGLIPLLICYKESVDQGNQIMSKRNVILFSVEDENRSWYLT 608
DB 501 TRPYISNVDITRDLASGLIGLILICKSRSLDRGIRQAADEQAVFAVEDEKNSWYIED 560
QY 609 NIQFLPNPAGVQLEDPEFQASNMIMH-----SINGVVFDSIQ--LSVCLHEAVYWYILSIG 662
DB 561 NIYKFCENPEYKRDPPKFEYSNIMSNFTLPAINGYVPESIPILGFCFDDTVQWHFCSVG 620
QY 663 AQTDLSVFEFSGYTFKHKMYEDTLTFPFGSETVFMSENGMLILGCHNSDFRNRGMT 722
DB 621 TONDILTIHFTGHSFIYGRKHEDTLTLFPMQGESVTVTMDNVGTWMLTTMNSNPRSKRLR 680

QY 723 ALLKVSSCDKNTGD-----YYEDS-----YEDISAYLL-- 751
DB 681 LRFRAKACIRNDDDDSYEIIYEPGSGTAMTTKKIHDSSEIEDENDADSDYODELALILGL 740
QY 751 -----SKNNAIEPRFSQ----- 764
DB 741 RSFRNSLNOEKDELNTALALEKDSFEIIPPSANRSLDSNSSSRSHVRLIAKNFAESLK 800
QY 764 -----NPPVLK----- 770
DB 801 TLHLLEAPAGSPLEHAGLDKNSALNPMAEHSSPYSEDPREDHPLSDVTGVSLLPFGTG 860
QY 770 -----RHQR-----EITRTIQSDQ-----EIT----- 788
DB 861 FKNRKPAAKHQRFQVGRGOAKHKFSQTRPPAKTTRLSQDNSSSSRMGPWEDIPSDLL 920
QY 788 ----- 788
DB 921 LQOKDPYKILNGEWHLYSEKSYEIIODANENKTVNKLPNSPONDSTRWGENIPFKNSHG 980
QY 788 ----- 788
DB 981 KQSGHPFLVTRRKPIQDRDRNRSLKEGLPLIRTRKKKEKPAYHVPLSPRSFHLPLR 1040
QY 788 -----DYDD----- 792
DB 1041 GEVNASFSDRRHNSLLHASNETSLIDLQTFPSMNLASLAPDHQTSFNDTTSQT 1100
QY 792 -----TISEMKKEDDIYED-----ENQSP----- 814
DB 1101 SSPDLTYPSPEEHYQIFPIQSDDPHSTHSTAPSNSRSPDPHSTHSTAPSNSRSPDPHST 1160
QY 814 ----- 814
DB 1161 NYDLRNRAIPTDVSQIFPSLELVWQTATSLDLSQPSISPDLGOMALSPDQGESLSPDL 1220
QY 814 ----- 814
DB 1221 GQTSLSPLDQESLSPDLGQTALSPDPQESLSPDLGQTALSPDPQESLSPDLGQTALS 1280
QY 814 ----- 814
DB 1281 PDPGQESLSPDLGQTSLSPLDQESLSPDLGQTALSPDPQESLSPDLGQTALSPDPQ 1340
QY 814 ----- 814
DB 1341 SLSPDLGQTSLSPLDQESLSPDLGQTALSPDPQESLSPDLGQTSLSPLDQESLSPDL 1400
QY 814 ----- 814
DB 1401 GQTALSPDLQESLSPDLGQTPLSPLDQESLSPDLGQTLQTSPLDLNQTSHTSSESS 1460
QY 814 ----- 814
DB 1461 QSLPLPEFGQTFPNADIGQMPSPDPSTLNTFTIPEEFNPLVVGLSRDDGDYIEIIPRQ 1520
QY 814 -----RSFOKTRHYFIAVER 830
DB 1521 KEESSEEDYGEFEFVAYNDPYQTLRTDINSSRNPDNTIAWYLRSTNGNRKYYIAAEET 1580
QY 831 LMDYGMSSPHVLRNRAQSGS---VPQ---FKKVFOEFTDGSFTQPLRGELNEHLGLL 884
DB 1581 SWDYS-----KFVQSDVDVYVPEDYVKKVFRKYLDSFTFKLDPQGEYEHLGIL 1631
QY 885 GPYIRAEVEDNIMVTFRQASRPYSFYSLSISYE-----EDQROGAEPKKNFVKPNE 936
DB 1632 GPVIRAEVDDVYQVRFKNAASRPYSLHAAGLSYEKSSSECKTYEDDSPEMFKEDNAIQPNK 1691
QY 937 TKTYFWKVOHMAPTKDEDCAMAYESVDVLEKDVHSGGLIGPLLVCHTNTLNPAGHGV 996
DB 1692 TTYVWHAATTRSGPENPGSACRAMAYSAVNPEKDIHSGILIGPLICKRGTLDKETNMPV 1751

```

QY 997 TVQEFALFTIFDETSKSWYFTEENMERNCRAPCNQIMEDPTFKENYRFHAINGYIMDTLPG 1056
Db 1752 DMREFVLLFMVFEDEKSKSWYKPKPTRSWRRASS-----EYKNSHEFHAINGMIIYN-LPG 1804
QY 1057 LVMAQDQIRIRYLLSMGSNENIHSFSGHFTVRKKEEYKALNLYPGVEFTEVEMLP 1116
Db 1805 LRMEOEWVRHLHLNLGSGRDHVHFHGQTLLENGTQOHLGWPPLPGSKTLEMKAS 1864
QY 1117 KAGIWRVECLIGEHLHAGMSTLFLVSNKQPTPLGASGHTRDFQITASGYGOWAPKLA 1176
Db 1865 KPGWMLDTEVGEIQRAGMQPFLIVDRECKMPMGLSTGLIADSQIQASEFGWEPKLA 1924
QY 1177 RLHYSGSINAW-----STK-EPEFSWIKVDLLAPMIIHGIKTQGAROKFSSLYISQFIIMY 1230
Db 1925 RLNNGSYNAMIAEKLTSEFNEPPIQVDMQKEVLLTGITQGAHYLKPPTTEFCVAY 1984
QY 1231 SLDGKKWQTYRGNSTGTLMVFEGNVDSGKHNIFNPPIIARIRLHPHYIRSTLRME 1290
Db 1985 SLDRKNWRIRKGNSTRNVMYFGNSDASTIKENOIDPPVAVARIRISPTGSKPALRLE 2044
QY 1291 LMGCDLNSCMLPGMESKALSDAQITASSYFTNMFAT-WSPSKARLHLOGRSNAMPQVN 1349
Db 2045 LOGCEVNGCSTPLGMESKIKENKQITASSFEKSKSWGNWYEPFLARLNAQGRVANAWQAKAN 2104
QY 1350 NPKEWLOVDFOKTMKVTGVTQGVKSLTSMYVKEFLISSQDGHQWTLFQNGKV--KV 1407
Db 2105 NNNQWLQIDLKIKKITAIVTQCKSLSEMYKSYTIHYSDOGTDWKPYRKSSMWDKI 2164
QY 1408 FQGNQDSFTPVNSLDPLRLRYLRHPQSVWQIALRMEVYLGCEAODLY 1457
Db 2165 FEGNNNVRGHVKNFNPPIISRIRIIPKTNOSIALRLLEFGC--DMY 2211

```

```

RESULT 6
CERU_HUMAN
ID CERU_HUMAN STANDARD; PRT; 1065 AA.
AC P00450; Q14063;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CERULOPLASMIN PRECURSOR (EC 1.16.3.1) (FERROXIDASE).
GN CP.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
SEQUENCE FROM N.A.
MEDLINE; 86259737.
RA KOSCHINSKY M.L., FUNK W.D., VAN OOST B.A., MCGILLIVRAY R.T.A.;
RT "Complete cDNA sequence of human preceruloplasmin.";
RL PROC. NATL. ACAD. SCI. U.S.A. 83:5086-5090(1986).
RN [2]
RP SEQUENCE OF 1-40; 549-599; 784-829 AND 919-952 FROM N.A.
RX MEDLINE; 86275241.
RA MERCER J.F.B., GRIMES A.;
RT "Isolation of a human ceruloplasmin cDNA clone that includes the N-
terminal leader sequence.";
RL FEBS LETT. 203:185-190(1986).
RN [3]
RP SEQUENCE OF 218-1065 FROM N.A.
RX MEDLINE; 86205876.
RA YANG F., NAYLOR S.L., LUM J.B., CUTSHAW S., MCCOMBS J.L.,
RA NABERHAUS K.H., MCGILL J.R., ADRIAN G.S., MOORE C.M., BARNETT D.R.,
RA BOWMAN B.H.;
RT "Characterization, mapping, and expression of the human ceruloplasmin
gene.";
RL PROC. NATL. ACAD. SCI. U.S.A. 83:3257-3261(1986).
RN [4]
RP SEQUENCE OF 20-1065.
RX MEDLINE; 84119493.
RA TAKAHASHI N., ORTEL T.L., PUTNAM F.W.;
RT "Single-chain structure of human ceruloplasmin: the complete amino
acid sequence of the whole molecule.";

```

```

RL PROC. NATL. ACAD. SCI. U.S.A. 81:390-394(1984).
RN [5]
RX SEQUENCE OF 158-333; 518-724 AND 858-1065.
RA MEDLINE; 83117800.
RA TAKAHASHI N., BAUMAN R.A., ORTEL T.L., DWULET F.E., WANG C.-C.,
RA PUTNAM F.W.;
RT "Internal triplication in the structure of human ceruloplasmin.";
RL PROC. NATL. ACAD. SCI. U.S.A. 80:115-119(1983).
RN [6]
RP SEQUENCE OF 501-905.
RX MEDLINE; 81199407.
RA DWULET F.E., PUTNAM F.W.;
RT "Complete amino acid sequence of a 50,000-dalton fragment of human
ceruloplasmin.";
RL PROC. NATL. ACAD. SCI. U.S.A. 78:790-794(1981).
RN [7]
RP SEQUENCE OF 907-1065.
RX MEDLINE; 80137543.
RA KINGSTON I.B., KINGSTON B.L., PUTNAM F.W.;
RT "Primary structure of a histidine-rich proteolytic fragment of human
ceruloplasmin. I. Amino acid sequence of the cyanogen bromide
peptides.";
RL J. BIOL. CHEM. 255:2878-2885(1980).
RN [8]
RP SEQUENCE OF 907-1065.
RX MEDLINE; 80137544.
RA KINGSTON I.B., KINGSTON B.L., PUTNAM F.W.;
RT "Primary structure of a histidine-rich proteolytic fragment of human
ceruloplasmin. II. Amino acid sequence of the tryptic peptides.";
RL J. BIOL. CHEM. 255:2886-2896(1980).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS).
RA ZAITSEVA I., ZAITSEV V., CARD G., MOSHKOV K., BAX B., RALPH A.,
RA LINDLEY P.;
RT "The X-ray structure of human serum ceruloplasmin at 3.1 A: nature of
the copper centres.";
RL J. BIOL. INORG. CHEM. 1:15-23(1996).
CC -I- FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING (6-7 ATOMS PER
MOLECULE) GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS
ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT
AND HOMEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.
CC -I- CATALYTIC ACTIVITY: 4 FE(2+) + 4 H(+) + O(2) = 4 FE(3+) + 2 H(2)O.
CC -I- COFACTOR: BINDS 6 CU-IONS PER MOLECULE.
CC -I- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH
CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE
2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR.
CC -I- TISSUE SPECIFICITY: SYNTHESIZED IN LIVER AND SECRETED INTO THE
PLASMA.
CC -I- DISEASE: CERULOPLASMIN IS DEFICIENT IN WILSON'S DISEASE.
CC -I- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
2 PLASTOCYANIN-LIKE REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M13699; G180256; -.
DR EMBL; D00025; D1000445; -.
DR EMBL; X04135; G30174; -.
DR EMBL; X04136; E3928; -.
DR EMBL; X04137; E3929; -.
DR EMBL; X04138; E3930; -.
DR EMBL; M13536; G180249; -.
DR PIR; A25443; KUH0.
DR PIR; A24165; A24165.
DR PDB; 1KCM; 12-FEB-97.
DR SWISS-2DPAGE; P00450; HUMAN.
DR MIM; 117700; -.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.

```

PROSITE: PS00080; MULTICOPPER_OXIDASE2; 1.
DR PFAM: PF00394; Cu-oxidase; 3.
KW OXIDOREDUCTASE; COPPER; METAL-BINDING; GLYCOPROTEIN; PLASMA; REPEAT;
KW SIGNAL; POLYMORPHISM; 3D-STRUCTURE.
ET SIGNAL 1 19
ET CHAIN 20 1065 CERULOPLASMIN.
ET DOMAIN 20 357 F5/8 TYPE A 1.
ET DOMAIN 20 200 PLASTOCYANIN-LIKE 1.
ET DOMAIN 209 357 PLASTOCYANIN-LIKE 2.
ET DOMAIN 370 718 F5/8 TYPE A 2.
ET DOMAIN 370 560 PLASTOCYANIN-LIKE 3.
ET DOMAIN 570 718 PLASTOCYANIN-LIKE 4.
ET DOMAIN 730 1061 F5/8 TYPE A 3.
ET DOMAIN 730 900 PLASTOCYANIN-LIKE 5.
ET DOMAIN 908 1061 PLASTOCYANIN-LIKE 6.
ET CARBOHYD 138 138
ET CARBOHYD 358 358
ET CARBOHYD 397 397
ET CARBOHYD 762 762
ET DISULFID 174 200
ET DISULFID 276 357
ET DISULFID 534 560
ET DISULFID 637 718
ET DISULFID 874 900
ET METAL 120 120
ET METAL 122 122
ET METAL 180 180
ET METAL 182 182
ET METAL 994 994
ET METAL 997 997
ET METAL 999 999
ET METAL 1039 1039
ET METAL 1040 1040
ET METAL 1041 1041
ET METAL 1045 1045
ET METAL 1050 1050
ET VARIANT 79 79
ET VARIANT 449 449
ET VARIANT 1060 1060
ET CONFLICT 1065 AA; 122205 MW; D127894A CRC32;
SO SEQUENCE

Query Match 22.3%; Score 1741; DB 1; Length 1065;
Best Local Similarity 32.6%; Pred. No. 8.5e-105;
Matches 389; Conservative 196; Mismatches 416; Indels 194; Gaps 23;

5 LSTCFELCLLRFCSATRRYYLGAVELSDYMSDGE--LPVDARFP RPVPKSPFNT 61
6 LGIFLFLCSTP-AWAKEKHYYIGIETWYD-ASDHGEKLLISVDTEHSNIYLONGDRI 63
62 SVVYKKTLLVEFTDHLFNIAKRRPRMGLGPTQAEVYDVTITLKNMASHPVSLHVG 121
64 GRLYKKALYQYTDTEFRTTEKRPVWLGLPITKAETGDKYVYHLKSLASRPYTHSHG 123

122 VSYMKASGAEYDDQTSQREKEDKYPGGSHTYVVOYLKENGPMASDPLCLTYSLSHV 181
124 ITYYKEHEGAIPPNDITDFQRADDKYYPGEQYTYMLATEEQSPGEDGNCVTRITHSHI 183
162 DLVKNLNSGLIGALLVCREGLAKEKQTL-HKFIILFAVDEGKSWHSE-----TKNS 234
184 DAPKDIASGLIGPLICKDLSLDEKEKHIDREFVYMFVSDENESWYLEDNITKTCSEP 243
235 LMQDRDASARAMPKMTVNGVYVNRSLPGLIGCHKRSYVWHVIGMTTPREVHSIFLEGHT 294
244 EKVDKDNEDFQESNRKMSVNGYTFGSLPGLSMCAEDRYKWLFLGMEVNDVHAAFFHQQA 303
295 FLVNRHQASLEISPTFLTAQTLMDLGOLFCHTSSHOHDMGEAVYKVDSCPEEPQL 354
304 LTNKNYRIDTINLFPATLLFDAYVVAQNPGEWMLSCQNLNLKAGLQAFQVOQEC----- 358
355 RKNNEAEADYDDDLTIDSEMDVVRFDNNSPSFIQIRSVAKKHPTVWHYIAAEEDMDY 414
358 ---NKSXSKD-----NIRGKHVRH-----YIIAAEEITWNY 385

415 APL-----VLAPDRSYKSOYLNGPQIRGRKKYVRMAVYTDTEF---KTREAIQ 462
386 APSGIDIFTKENLAPGSDS--AVFEQGTTRIGSSYKLVYRETTDASFTNKRERGPEE 443
463 HESGILGPLLYGEVDTLIIIFKNQASRPYNIYPHGI-----TDVRPLY---SRRLP 511
444 EHLGILGPVIAEAGDTIRVTFHNKGAAYPLSTIEPIGVRFNKNNECTYSSPNYNQSRSPV 503
512 KGVKHLKDFPILPGEIIFKMYVYVEDGPTKSDPRCLTRYSSFFNMERDLASGLIGPLL 571
504 PSASH-----VAPTEFTYEWYVKEVGPTNADPYCLAKMYSAVDPTKDIFTGLIGPMK 558
572 ICYKESVDQRGNQMSDKRNVLFSVFEDENRSWYLTENIQRFLEPNPAGVQLEDPEEQASN 631
559 ICKGSLHANGRQKDYDKEFYLFPTVFEDENESLLEDNIRMFTTAPDQVDKEDPEQESN 618
632 IMHSINGVYFDSLO-LSVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKHKMYEDTLTLF 690
619 KMSMNGFMYGNOGLTMCCKGDSVYWYILFSAGNEADVHGIFYSGNTYLMRGERDITANLF 678
691 PFSGETVMSMENGLMILGCHNSDFRNKGMATALLKVVSSCDKNTGDIYEDSYEDISAYLL 750
679 PQSLTLHMPDTEGTFNVECLTIDHTYTGMMKQKYTVNQCRQS-----EDS----- 726
751 SKNNAIEPRSFQNPVYLKRHQREITRTTLOSQDEIDYDDTISVEMKKEDFDIYDEDEN 810
726 ----- 726
811 QSPRSFOKTRHYFIAAVERLWDYGMSSP-----HVLNRQAQSGSV-----P 853
726 ---TFYLGERTYIAAVEWMDY---SPQREWEKELHHLQEQVNSNAFLDKGEFYIGS 777
854 QFKVYVQEFQFTDGSFTQPLYRGELENEHLGLPYIRAEVEDNIMVTFRNOASRPYSFYSS 913
778 KYKVVYRQYTDSTFRVPERKABEEHLGILGPQIHADVGDKVKIIFKNMATRPSYIHA- 837
914 LISYEDQROGAEPKKNFVK--ENETKYFWKVQHHMAPTDEFDCKAWAYFSDVLEK 970
837 -----HGVOESSTVTPPLPGETLTYVWKIPERSGAGTEDSACIPWAYYSYVDQVK 887

971 DVHSGLIGPLLVCHNTLNPAGROVTVQEFALFETTFIDETKSWYFTEMERNCRAPCNI 1030
888 DLVSGLIGPLIVCRPRLVKFNPRKL--EFALFLVFEDENESWYLDNITKYSDBPEKV 945
1031 QMEDPTFKENYRFAINGYIMDTPLGLVMAODORIRWYLLSMGSNENIHSIHSGHYFTV 1090
946 NKDEEFIESNKNMAINGRMFGNLQGLTMHVGDENVWYLMGMGNEIDLHTVHFHGSFYQ 1005
1091 RKKEEYKALYNLYPGVEFTEVEMLPKAGIWRVECLIGEHLHAGNSTLFLVYSNK 1145
1006 KHRGVYSSDVFDFIPGTYQTLMEPRTPGIWLHCHVTVDHIAHAGETTYTVLQNE 1060

RESULT 7
CERU_RAT
ID CERU_RAT STANDARD; PRT; 1059 AA.
AC P13635; 064719;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CERULOPLASMIN PRECURSOR (EC 1.16.3.1) (FERROXIDASE).
GN CP.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LUNG, AND LIVER;
RX MEDLINE; 90237081.
RA FLEMING R.E., GITLIN J.D.;
RT "Primary structure of rat ceruloplasmin and analysis of
tissue-specific gene expression during development.";

RL J. BIOL. CHEM. 265:7701-7707(1990).
RN [2]
RP SEQUENCE OF 257-294; 571-612 AND 823-892 FROM N.A.
RC TISSUE=LIVER.
RX MEDLINE; 87137545.
RA ALDRED A.R., GRIMES A., SCHREIBER G., MERCER J.F.B.;
RT "Rat ceruloplasmin. Molecular cloning and gene expression in liver,
choroid plexus, yolk sac, placenta, and testis.";
RL J. BIOL. CHEM. 262:2875-2878(1987).
CC -1- FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING (6-7 ATOMS PER
MOLECULE) GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS
ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT
AND HOMEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.
CC -1- FUNCTION: MAY ALSO PLAY A ROLE IN FETAL LUNG DEVELOPMENT OR
PULMONARY ANTIOXIDANT DEFENSE.
CC -1- CATALYTIC ACTIVITY: 4 FE(2+) + 4 H(+) + O(2) = 4 FE(3+) + 2 H(2)O.
CC -1- COFACTOR: BINDS 6 CU-IONS PER MOLECULE.
CC -1- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH
CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE
2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR.
CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN LIVER AND SECRETED INTO THE
PLASMA. ALSO CHOROID PLEXUS, YOLK SAC, PLACENTA, AND TESTIS; NOT
IN STOMACH AND SMALL INTESTINE. FETAL LUNG AND LIVER.
CC -1- INDUCTION: BY INFLAMMATION.
CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
2 PLASTOCYANIN-LIKE REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L33869; G499669; -
DR EMBL; M80529; G203421; -
DR EMBL; J02670; G203416; ALT_SEQ.
DR EMBL; M14102; G554429; -
DR PIR; A35210; A35210.
DR PIR; A29564; A29564.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
DR PFAM; PF00394; Cu-oxidase; 3.
DR HSSP; P00450; 1KCV.
OXIDOREDUCTASE; COPPER; METAL-BINDING; GLYCOPROTEIN; PLASMA; REPEAT;
SIGNAL.
CC -----
CC SIGNAL. 1 19 PROBABLE.
FT CHAIN 20 1059 CERULOPLASMIN.
FT DOMAIN 20 356 F5/8 TYPE A 1.
FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.
FT DOMAIN 208 354 PLASTOCYANIN-LIKE 2.
FT DOMAIN 369 712 F5/8 TYPE A 2.
FT DOMAIN 369 554 PLASTOCYANIN-LIKE 3.
FT DOMAIN 564 710 PLASTOCYANIN-LIKE 4.
FT DOMAIN 724 1055 F5/8 TYPE A 3.
FT DOMAIN 724 894 PLASTOCYANIN-LIKE 5.
FT DOMAIN 902 1051 PLASTOCYANIN-LIKE 6.
FT DISULFID 173 199 BY SIMILARITY.
FT DISULFID 275 356 BY SIMILARITY.
FT DISULFID 528 554 BY SIMILARITY.
FT DISULFID 631 712 BY SIMILARITY.
FT DISULFID 868 894 BY SIMILARITY.
FT METAL 120 120 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 122 122 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 179 179 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 181 181 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 988 988 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 991 991 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 993 993 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1033 1033 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1034 1034 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 1035 1035 COPPER (TYPE 3) (BY SIMILARITY).

FT METAL 1039 1039 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 1044 1044 COPPER (TYPE 1) (BY SIMILARITY).
FT CARBOHYD 138 138 POTENTIAL.
FT CARBOHYD 226 226 POTENTIAL.
FT CARBOHYD 396 396 POTENTIAL.
FT CARBOHYD 582 582 POTENTIAL.
FT CARBOHYD 756 756 POTENTIAL.
FT CARBOHYD 920 920 POTENTIAL.
FT CONFLICT 271 271 G -> A (IN REF. 2).
FT CONFLICT 604 605 ED -> DN (IN REF. 2).
FT CONFLICT 823 823 T -> S (IN REF. 2).
FT CONFLICT 833 833 V -> L (IN REF. 2).
FT CONFLICT 868 868 C -> V (IN REF. 2).
FT CONFLICT 891 891 L -> R (IN REF. 2).
SQ SEQUENCE 1059 AA; 120840 MW; 8E9F8FAD CRC32;

Query Match 21.7%; Score 1694; DB 1; Length 1059;
Best Local Similarity 32.5%; Pred. No. 9.2e-102;
Matches 386; Conservative 188; Mismatches 428; Indels 186; Gaps 23;

OY 5 LSTCFELCLRFCSATRRYYLGAVELSDWYMQ-SDLGEL-PVDARFPPRPVKSPFNTS 62
|| || : : : || | | | : : |
Db 6 LSALEFL-HSLAWREKHYYIGITEAVWDYASGESEKELISVDEQSNFYLRNGPDRIG 64

OY 63 VVYKTLFEFTDHLFNIAKPRPPMGLGPTIOAEVYDTVTILKMAASHPVSLHAGV 122
|| | : : | | : : | | | | | : : | | | | : : | | |
Db 65 RYKKALEYETDGTFTKIDKPAMGLFGLPVKAEVGDKVSHVKNFASRPYTHAGV 124

OY 123 SYWKASGAEYDDQTSOREKEDKYPPGSHTYWQVLEKNGPMASDPLCLTYLSHVD 182
: || : || | : : : || : || | : : | : | : | : |
Db 125 TYTKANGAIPDNTDFORADKLFPGQOYLVLRA-NEPSPGEDSDNCVTRIYSHVD 183

OY 183 LVKDLNSGLIGALLVREGSLAKEKQTL-HKFIILFAVDEGKSHSETKNSLM----- 237
|| : || | | : : || : || | : : : || : || | : |
Db 184 APKDIASGLIGPLILCKKGLSKHKEENIDQEFVIMFSVDENLSMTLEDNIKTCSEPE 243

OY 237 -QDRDAASARAMPKMTVNGYVNRSLPGLIGCHRKSYVHWIGMTTPEVHSIFEGHTE 295
| : | : : : || | | | | : | : | : | : | : |
Db 244 KVDKDNEDEQESNRMYSSINGYTFGLSLGSMCAEDRYKWLFGMGNEVDVHSELFGQAL 303

OY 296 LVNRHQAASLEISPTFELTAQTLMDIGQFLFCHISSHQHDMGMAVYKVDSCPEEPQLR 355
: | : : : | : : : : : : : : : : : : : : : :
Db 304 TSKNYHTDILNLPATLIDVSMVAQNGVWMLSCQNLHLKAGLQAFQVYRDC----- 357

OY 356 MKNNEAEADYDDDLTJSEMDEVRFDDNDSPSFIQIRSVAKKHPKTYVHYIAEEDMDYA 415
| : : || : | : : : : : : : : : : : : : : :
Db 357 --NKPS--DDDIQDRHV-----RH-----YIIAEETIDYA 385

OY 416 P-----LVLPDDRYSYKQYLNNGPQIRGRYKVKVRFMAYTDEF---KTRBAI 461
| | : | : | : | : | : | : | : | : | : | : |
Db 386 PSGLDTFTGENFTSLGSDSRVFEQ---GATRIGGSYKLVLYREYTDSDFTNRKERGPD 441

OY 462 QHESGILGPLLYGEVGTLLLIIFKNQASRPYNIYPHGITDVR---PLYSRRLPKGVKHL 517
: || || : : : : : : : : : : : : : : : : : :
Db 442 EHLGILGPVIAEAVGDIIRVTFHNKGOPPLSIQPMGVHFTKENEGTYG---PDGRSK 498

OY 518 KDEPILPGEIIFKYKWTIVVEDGPTKSDPRCLTRYSSFYVMERDLASGLIGPLLCYKES 577
: : : | : | : | : | : | : | : | : | : | : | :
Db 499 QASHVAPKETFTYEWTPREMGPTYADPCLSKMYSGVDLTKDIFGLIGPMKICKKGS 558

OY 578 VDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRLPMPAGVOLLEDPFQASNIMHSTIN 637
: | : | : : || | | : : || : | : | : | : | : |
Db 559 LLADGRQKDYDKEFFYLFAVFDENESLLLDNIRMTTAPENVDKEDPFQESNKMHSMN 618

OY 638 GYVEDSLQ-LSVCLHEVAYWYILSIGAQDTFLSVFSGYTFKHKMYEDTLTLPFSGET 696
| : : : : | : | : | : | : | : | : | : | : | :
Db 619 GFMYGNLPGLNMLGESIYWYLFSGAGNEADVHGIFYSGNTYLSKGERDTANLFPKSLT 678

OY 697 VMSMENPGIILGCHNSDFRNRGMTALLKYSSCDKNTGDIYEDYSAYLLSKNNAI 756
: | : : | : : : | : | : | : | : | : | : : :
Db 679 LIMTPDTEGSDVECLTTDHYTGMMKQKYTVNQC-----KGQFEDVTLY----- 723

QY 757 EPRFSQNPVYLKRHOREITRTTLOSQDEIDYDDTISVENMKKEDFDIYDEDEENQSPRSF 816
Db 723 ----- 723
QY 817 OKTRHYFLAVERLMDYGMSSPHVLNRNRAOSGSV-----POFKKVVFOEF 863
Db 723 -QERTYTLAAVEVEMDYSPPSDWEMELHLEQONVNAFLDKKEFFIGSKYKVVYREF 781
QY 864 TDGSEFTQPLRGELNHLGLGPIYIRAEVEDNIMVTFRNQASRPYSFYSLISYEEDQR 923
Db 782 TDSTFREQYKRAEEHGLMGLPLIHADYKAKVVFKNMATRPSIHA-----H 831
QY 924 GAEPKKNFVK--PNETKTYFWKVQHHMAPTKDEFDCAMAYFSDVDLEKDVHSLIGPL 980
Db 832 GVKTKSSTVAPTLPGEVRTYIQIPERSGACTEDSPCIPMAYYSTVDKVKDLYSLIGPL 891
QY 981 LVC---HTNTLNPAHGRQVYVQEFALFTTFDETKSWYFTEENMRNCRAPNIOMEDPTF 1037
Db 892 IVCRKSYKVENPK-----KMEFSLFLVEDENESWYLDNINITYPDHPEKDNKDNKEEF 946
QY 1038 KENYRFHAIINGYIMDTLPGLVMAQDQIRIRWYLLSMGSNENIHSFGHVFTVRKKEEK 1097
Db 947 IESNMHAIKMGKFMGNLQGLTMHVGDENVNMYMANGNEIDLHTVHFHGHSGFYKRGHIS 1006
QY 1098 MALYNLPYGVETVEMLPKAGIWRVECLIGELHAGMSTLEFLVYSNK 1145
Db 1007 SDVFDEFPGTYQTLEMFPGTPGTWLLHCHVTDHIHAGMVTYTVLPNQ 1054

RESULT 8
CERU_MOUSE STANDARD; PRT: 1062 AA.
ID CERU_MOUSE 061147;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CERULOPLASMIN PRECURSOR (EC 1.16.3.1) (FERROXIDASE).
GN CP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA KLOMP L.W.J., FARHANGRAZI Z.S., CHOI D.W., GITLIN J.D.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP TISSUE SPECIFICITY.
RX MEDLINE; 96294736.
RA KLOMP L.W.J., FARHANGRAZI Z.S., DUGAN L.L., GITLIN J.D.;
"Ceruloplasmin gene expression in the murine central nervous system.";
J. CLIN. INVEST. 98:207-215(1996).
-1- FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING (6-7 ATOMS PER
MOLECULE) GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS
ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT
AND HOMEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.
-1- CATALYTIC ACTIVITY: 4 FE(2+) + 4 H(+) + O(2) -> 4 FE(3+) + 2 H(2)O.
-1- COFACTOR: BINDS 6 CU-IONS PER MOLECULE.
-1- COFACTOR: THIS PROTEIN BELONGS TO THE MULTICOPPER OXIDASES WHICH
CONTAIN THREE DISTINCT CU CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE
2 OR NORMAL, AND TYPE 3 OR COUPLED BINUCLEAR.
-1- TISSUE SPECIFICITY: MANY TISSUES, INCLUDING LIVER, EYE AND BRAIN.
-1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
2 PLASTOCYANIN-LIKE REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U49430; G1224108; -.

DR MGD; MG1:88476; CP.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
DR PFAM; PF00394; Cu-oxidase; 3.
DR HSSP; P00450; 1KCM.
KW OXIDOREDUCTASE; COPPER; METAL-BINDING; GLYCOPROTEIN; PLASMA; REPEAT;
KW SIGNAL.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 1062 CERULOPLASMIN.
FT DOMAIN 20 356 F5/8 TYPE A 1.
FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.
FT DOMAIN 208 356 PLASTOCYANIN-LIKE 2.
FT DOMAIN 369 713 F5/8 TYPE A 2.
FT DOMAIN 369 555 PLASTOCYANIN-LIKE 3.
FT DOMAIN 565 713 PLASTOCYANIN-LIKE 4.
FT DOMAIN 725 1057 F5/8 TYPE A 3.
FT DOMAIN 725 896 PLASTOCYANIN-LIKE 5.
FT DOMAIN 904 1057 PLASTOCYANIN-LIKE 6.
FT DISULFID 173 199 BY SIMILARITY.
FT DISULFID 275 356 BY SIMILARITY.
FT DISULFID 529 555 BY SIMILARITY.
FT DISULFID 632 713 BY SIMILARITY.
FT DISULFID 870 896 BY SIMILARITY.
FT METAL 120 120 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 122 122 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 179 179 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 181 181 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 990 990 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 993 993 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 995 995 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1035 1035 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1036 1036 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 1037 1037 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 1041 1041 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1046 1046 COPPER (TYPE 1) (BY SIMILARITY).
FT CARBOHYD 138 138 POTENTIAL.
FT CARBOHYD 226 226 POTENTIAL.
FT CARBOHYD 583 583 POTENTIAL.
FT CARBOHYD 625 625 POTENTIAL.
FT CARBOHYD 757 757 POTENTIAL.
FT CARBOHYD 922 922 POTENTIAL.
SQ SEQUENCE 1062 AA; 121159 MW; 946E9EE7 CRC32;

Query Match 21.28; Score 1651; DB 1; Length 1062;
Best Local Similarity 32.38; Pred. No. 5.5e-99;
Matches 382; Conservative 196; Mismatches 443; Indels 160; Gaps 25;

QY 10 FLCLLRFCF-----SATRRVYLGAVELSWDYMQ--SDIGELPVDAFPKSPFPN 60
Db 3 FLLSTFIFLYSSLALARDKHVFIGITEAVWDYASGTEKKLISVDTEQSNFYLONGPDR 62
QY 61 TSVYKKTLFEFTDHLFNIAKPRPPWVGILPTIQAEVYDTVVITLKNMASHPVSLHAV 120
Db 63 IGRKYKALYFEYTDGTFSKRTIDKPWWLGLGVKAEVEDKVVVHLKNLASRIYTFHAF 122
QY 121 GVSYWKASGAEYDDQTSOREKEDDKVPGSHTYVWQVYKENGPMASDPLCTYSYLSH 180
Db 123 GVTYTKYEGAVYPDNTDFQRADDKVLPQQYVYVLAH-NEPSGEGDSNCVTRIYHSH 181
QY 181 VDLVKDLSGLIGALLVCRGSLAKEKQTQL-HKFIILFAVFEDEGKSWHSETKNSLM--- 237
Db 182 VDAPKDIASGLIGPLILCKKGSGLYKEKEKNIDQEFVLMFSVVDENLSWYLEDNIKTFCS 241
QY 237 ---QDRDAASARAMPKMHITVNGYVNRSLPGLIGCHRSYVYHVGIMGTPEVHSIFLEGH 293
Db 242 PEKVDKDNEDEQESNRMYSSINGTYTGLSLPGLSMCAADRVKRWYVLFMGNEVDVHSAFFHQ 301
QY 294 TFLVRNHQASLEISPIFTFLAQTLLMDLQGLFLFCHISSHQHDMGEAVYKVDSCPEEPQ 353
Db 302 ALTSRNYQTDIINLFPATLIDAYVVAQNPVWMLSCQNLNHLKAGLQAFQVQDC----- 357
QY 354 LRMKNNEADYDDDLTJSEMDVVRFFDDNSPSSIQIRSVAKKHKPKTWVHYIAAEEEDWD 413

316 KINAWTAQSNASAKEMLOVDLGTOKKVTGIIITOGARDFGHIQYVASYKVAHSDGQVQWTVY 375
QY 1400 FQNGKVKVFEQGNDSFTPVNSLDPELLTRYLRHQSWVHQIALRMEVLGC 1451
DB 376 EEOGTSKVFQGNLDNNSHKKNIFEKPFMARYVRYVLPDSWHRITLRLLELLGC 427

RESULT 10
MEGM_MOUSE STANDARD; PRT; 463 AA.
AC P21956; P97800; 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE MILK FAT GLOBULE-EGF FACTOR 8 PRECURSOR (MEG-E8) (MEGM) (SPERM SURFACE
DE PROTEIN SP47) (MP47).
GN MFE8.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
N [1]
N SEQUENCE FROM N.A., AND SEQUENCE OF 23-35.
N TISSUE=MAMMARY GLAND;
RX MEDLINE: 91046008.
RA STUBBS J.D., LEKUTIS C., SINGER K.L., BUI A., YUZUKI D.,
RA SRINIVASAN U., PARRY G.;
RT "cDNA cloning of a mouse mammary epithelial cell surface protein
RT reveals the existence of epidermal growth factor-like domains linked
RT to factor VII-like sequences.";
RL PROC. NATL. ACAD. SCI. U.S.A. 87:8417-8421(1990).
RN [2]
RP SEQUENCE OF 23-456 FROM N.A.
RC TISSUE=TESTIS;
RA ENSSLIN M.A.;
RL SUBMITTED (MAR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC BINDING PROTEIN.
CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC SPERMATOZOAN.
CC -1- DEVELOPMENTAL STAGE: MRNA EXPRESSION IS DETECTABLE IN MAMMARY
CC TISSUE FROM NONPREGNANT ANIMALS & MAXIMAL IN THE LACTATING GLAND.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M38337; G199143; -
CC EMBL: Y11684; E307035; -
CC PIR: A36479; A36479.
CC MGD: MGI:102768; MFE8.
DR PROSITE: PS00022; EGF_1; 2.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01285; FA58C_1; 2.
DR PROSITE: PS01286; FA58C_2; 2.
DR PFAM: PF00008; EGF; 2.
DR PFAM: PF00754; F5_F8_type_C; 2.
DR HSSP: P00740; IIXA.
KW SIGNAL; GLYCOPROTEIN; REPEAT; EGF-LIKE DOMAIN; MILK.
FT SIGNAL 1 22
FT CHAIN 23 463 MILK FAT GLOBULE-EGF FACTOR 8.
FT DOMAIN 24 61 EGF-LIKE 1.
FT DOMAIN 64 108 EGF-LIKE 2.
FT DOMAIN 148 303 F5/8 TYPE C 1.
FT DOMAIN 308 463 F5/8 TYPE C 2.
FT SITE 87 89 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 28 39 BY SIMILARITY.

FT DISULFID 33 49 BY SIMILARITY.
FT DISULFID 51 60 BY SIMILARITY.
FT DISULFID 68 79 BY SIMILARITY.
FT DISULFID 73 96 BY SIMILARITY.
FT DISULFID 98 107 BY SIMILARITY.
FT DISULFID 148 303 BY SIMILARITY.
FT DISULFID 290 294 BY SIMILARITY.
FT DISULFID 308 463 BY SIMILARITY.
FT CARBOHYD 61 61 POTENTIAL.
FT CARBOHYD 266 266 POTENTIAL.
FT CARBOHYD 316 316 POTENTIAL.
FT CARBOHYD 426 426 POTENTIAL.
FT CARBOHYD 30 30 S -> F (IN REF. 2).
FT CONFLICT 35 35 N -> D (IN AA SEQUENCE; REF. 1).
FT CONFLICT 55 55 F -> S (IN REF. 2).
FT CONFLICT 67 67 P -> G (IN REF. 2).
FT CONFLICT 110 147 ETNYNLDGEYMTTAVPTAVPTAPTPDLSNLSR ->
FT CONFLICT 168 168 G (IN REF. 2).
FT CONFLICT 196 168 Y -> S (IN REF. 2).
FT CONFLICT 309 196 H -> T (IN REF. 2).
FT CONFLICT 394 309 L -> S (IN REF. 2).
FT CONFLICT 395 395 VE -> AQ (IN REF. 2).
FT CONFLICT 410 410 V -> D (IN REF. 2).
SQ SEQUENCE 463 AA; 51465 MW; 254BAB08 CRC32;

Query Match 8.4%; Score 657; DB 1; Length 463;
Best Local Similarity 42.2%; Pred. No. 2.9e-35;
Matches 135; Conservative 63; Mismatches 110; Indels 12; Gaps 5;

QY 1143 SNKCOTPLGMSGHIRDFOITASGQY-----GOWAPKLARLHYSGSINAW--STKEPFS 1194
DB 145 ASRCSTQLMEGALADSQISASYVMGFGLRWGPRLARLRYRTGIVNMAHASNYSKDP 204
QY 1195 WIKVDLLAPMITHGIKTQGARQKFFSLYISQITIMYSLDGKKWQTYRGNSTGLMVFFGN 1254
DB 205 WIOYNLLRKMRSQVMTQGASAGRAEYLRKTYVAISLDGRKFEFIQDESGD-KEFLGN 263
QY 1255 VDSGKHNFPPIIARYIRLPHYISIRSLRMLMGLDSCSMPLGMSKAISDAQ 1314
DB 264 LDNNSLKVNMFPLEAQYIRLYPVSCHRCITLRFELLCGLPLEGLKNTIPDSQ 323
QY 1315 ITASSYFT--NMFA-TWSPSKARLHLQGRSNAWRPVNNPKEMLOVDFOKTKMVTGTTQ 1371
DB 324 MSASSSYKTWNLRAFGWYPHLGRLDNQGKINAWTAQSNASAKEMLOVDLGTORQVGTITQ 383
QY 1372 GVKSLTSMYKEFLISSQDGHQWTLFFQNGKVKVFOGNDSFTPVNSLDPELLTRYL 1431
DB 384 GARDFGHIQYVESYKVAHSDGQVQWTVYEQSSKVFQGNLDNNSHKKNIFEKPFMARYV 443
QY 1432 RIHPOSWVHQIALRMEVLGC 1451
DB 444 RVLPSWVHNRITLRLLELLGC 463

RESULT 11
MEGM_PIG STANDARD; PRT; 409 AA.
AC P79385; 01-NOV-1997 (REL. 35, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE MILK FAT GLOBULE-EGF FACTOR 8 (MEG-E8) (MEGM) (SPERM SURFACE PROTEIN
DE SP47) (PP47).
GN MFE8.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA ENSSLIN M.A.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

```

CC CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC BINDING PROTEIN.
CC CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC SPERMATOZOAN.
CC CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; Y11683; E1188617; -.
CC DR PROSITE; PS00022; EGF_1; 2.
CC DR PROSITE; PS01186; EGF_2; 2.
CC DR PROSITE; PS01285; FA58C_1; 2.
CC DR PROSITE; PS01286; FA58C_2; 2.
CC DR PFAM; PF00008; EGF; 2.
CC DR PFAM; PF00754; F5_F8_type_C; 2.
CC DR HSP; P00740; 11XA.
CC KW GLYCOPROTEIN; REPEAT; EGF-LIKE DOMAIN.
CC FT DOMAIN 2 41 EGF-LIKE 1.
CC FT DOMAIN 44 88 EGF-LIKE 2.
CC FT DOMAIN 91 247 F5/8 TYPE C.
CC FT DOMAIN 252 409 F5/8 TYPE C.
CC FT SITE 67 69 CELL ATTACHMENT SITE (POTENTIAL).
CC FT DISULFID 6 17 BY SIMILARITY.
CC FT DISULFID 11 29 BY SIMILARITY.
CC FT DISULFID 31 40 BY SIMILARITY.
CC FT DISULFID 91 247 BY SIMILARITY.
CC FT DISULFID 234 238 BY SIMILARITY.
CC FT DISULFID 252 409 BY SIMILARITY.
CC FT CARBOHYD 41 41 BY SIMILARITY.
CC FT CARBOHYD 372 372 POTENTIAL.
CC FT NON_TER 409 409
CC SQ SEQUENCE 409 AA; 45725 MW; 00F2880A CRC32;

```

Query Match 8.3%; Score 650; DB 1; Length 409;
Best Local Similarity 39.3%; Pred. No. 6.9e-35;
Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

[illegible]

RESULT 12
MEGM_BOVIN

```

ID      MEGM_BOVIN      STANDARD;      PRT;      427 AA.
AC      095114; 027959; P79344;
DT      01-NOV-1997 (REL. 35, CREATED)
DT      01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT      01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE      MILK FAT GLOBULE-EGF FACTOR 8 PRECURSOR (MEG-E8) (MGP57/53) (PAS-
DE      6/PAS-7 GLYCOPROTEIN) (MEGM) (SPERM SURFACE PROTEIN SP47) (BP47).
GN      MEGE8.
OS      BOS TAURUS (BOVINE).
OC      EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC      ARTIODACTYLIA; RUMINANTIA; PECORA; BOVIDEA; BOVIDAE; BOVINAE; BOS.
RN      [1]
RP      SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC      STRAIN-HOLSTEIN; TISSUE=MAMMARY GLAND;
RX      MEDLINE; 97008954.
RA      HVARREGAARD J., ANDERSEN M.H., BERGLUND L., RASMUSSEN J.T.,
RA      PETERSEN T.E.;
RT      "Characterization of glycoprotein PAS-6/7 from membranes of bovine
RT      milk fat globules."
RT      EUR. J. BIOCHEM. 240:628-636(1996).
RN      [2]
RP      SEQUENCE OF 18-427 FROM N.A.
RC      TISSUE=MAMMARY GLAND;
RX      MEDLINE; 96125736.
RA      AOKI N., KISHI M., TANIGUCHI Y., ADACHI T., NAKAMURA R.,
RA      MATSUDA T.;
RT      "Molecular cloning of glycoprotein antigens MGP57/53 recognized by
RT      monoclonal antibodies raised against bovine milk fat globule
RT      membrane."
RT      BIOCHIM. BIOPHYS. ACTA 1245:385-391(1995).
RN      [3]
RP      SEQUENCE OF 19-427 FROM N.A.
RC      TISSUE-TESTIS;
RA      ENSSLIN M.A.;
RL      SUBMITTED (MAR-1997) TO EMBL/GENBANK/DEJ DATA BANKS.
CC      -1- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
CC      MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
CC      BINDING PROTEIN.
CC      -1- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN.
CC      -1- ALTERNATIVE PRODUCTS: THERE EXIST 2 VARIANTS DUE TO ALTERNATIVE
CC      SPLICING; THE SHORTER FORM LACKS 53 AMINO ACIDS WITHIN THE F5/8
CC      TYPE C 1 DOMAIN.
CC      -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC      -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X91895; E204102; -
DR      EMBL; S80643; G1246081; -
DR      EMBL; Y11719; E307295; -
DR      PROSITE; PS00022; EGF_1; 2.
DR      PROSITE; PS01186; EGF_2; 2.
DR      PROSITE; PS01285; FA58C_1; 2.
DR      PROSITE; PS01286; FA58C_2; 2.
DR      PFAM; PF00008; EGF; 2.
DR      PFAM; PF00754; F5_F8_type_C; 2.
DR      HSP; P00740; IIXA.
DR      SIGNAL; GLYCOPROTEIN; MILK; REPEAT; EGF-LIKE DOMAIN;
KW      ALTERNATIVE SPLICING.
FT      SIGNAL      1      18
FT      CHAIN      19      427      MILK FAT GLOBULE-EGF FACTOR 8.
FT      DOMAIN      20      59      EGF-LIKE 1.
FT      DOMAIN      62      106      EGF-LIKE 2.
FT      DOMAIN      109     265      F5/8 TYPE C 1.
FT      DOMAIN      270     427      F5/8 TYPE C 2.
FT      SITE      85      87      CELL ATTACHMENT SITE (POTENTIAL).
FT      DISULFID      24      35      BY SIMILARITY.

```

FT	DISULFID	29	47	BY SIMILARITY.
FT	DISULFID	49	58	BY SIMILARITY.
FT	DISULFID	66	77	BY SIMILARITY.
FT	DISULFID	71	94	BY SIMILARITY.
FT	DISULFID	96	105	BY SIMILARITY.
FT	DISULFID	109	265	
FT	DISULFID	252	256	
FT	DISULFID	270	427	
FT	CARBOHYD	27	27	
FT	CARBOHYD	34	34	
FT	CARBOHYD	59	59	
FT	CARBOHYD	227	227	
FT	VARSPLIC	169	221	
FT	CONFLICT	19	19	
FT	CONFLICT	28	28	
FT	CONFLICT	53	53	
FT	CONFLICT	67	67	
FT	CONFLICT	149	149	
FT	CONFLICT	403	403	
FT	SEQUENCE	427	427	

Query Match	8.1%;	Score 635;	DB 1;	Length 427;
Best Local Similarity	37.9%;	Pred. No. 6.8e-34;		
Matches 135;	Conservative 66;	Mismatches 123;	Indels 32;	Gaps 7;

[illegible]

RESULT	13		
MFGM	HUMAN		
ID	MFGM_HUMAN	STANDARD:	PRT: 387 AA.
AC	Q08431:		
DT	01-OCT-1996	(REL. 34, CREATED)	
DT	01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)	
DT	15-JUL-1998	(REL. 36, LAST ANNOTATION UPDATE)	
DE	MILK FAT GLOBULE-EGF FACTOR 8 PRECURSOR (MFG-E8)	(HMFG)	(BREAST
DE	EPITHELIAL ANTIGEN BA46)	(MFGM).	
GN	MEGE8.		
OS	HOMO SAPIENS (HUMAN).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;		
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE-BREAST, AND BREAST CARCINOMA;		
RX	MEDLINE; 96213908.		
RA	COUTO J.R., TAYLOR M.R., GODWIN S.G., CERIANI R.L., PETERSON J.A.;		
RT	"Cloning and sequence analysis of human breast epithelial antigen		
RT	BA46 reveals an RGD cell adhesion sequence presented on an epidermal		

RT growth factor-like domain.";
RL DNA CELL BIOL. 15:281-286(1996).
RN [2]
RP SEQUENCE OF 170-387 FROM N.A.
RC TISSUE-MAMMARY GLAND;
RX MEDLINE; 91371351.
RA LARocca D., PETERSON J.A., URREA R., KUNIOYOSHI J., BISTRAN A.M.,
RA CERIANI R.L.;
RT "A Mr 46,000 human milk fat globule protein that is highly expressed
RT in human breast tumors contains factor VII-like domains.";
RL CANCER RES. 51:4994-4998(1991).
CC -I- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING.
CC -I- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -I- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES.
CC OVEREXPRESSED IN SEVERAL CARCINOMAS.
CC -I- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -I- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U58516; G1381162; -
CC

DR	PROSITE; PS00022; EGF_1; 1.
DR	PROSITE; PS01186; EGF_2; 1.
DR	PROSITE; PS01285; FA58C_1; 2.
DR	PROSITE; PS01286; FA58C_2; 2.
DR	PFAM; PF00008; EGF; 1.
DR	PFAM; PF00754; F5_F8_type_C; 2.
KW	SIGNAL; GLYCOPROTEIN; MILK; REPEAT; EGF-LIKE DOMAIN.
FT	SIGNAL 1 22 POTENTIAL.
FT	CHAIN 23 387 MILK FAT GLOBULE-EGF FACTOR 8.
FT	DOMAIN 23 67 EGF-LIKE.
FT	DOMAIN 70 225 F5/8 TYPE C 1.
FT	DOMAIN 230 387 F5/8 TYPE C 2.
FT	SITE 46 48 CELL ATTACHMENT SITE (POTENTIAL).
FT	DISULFID 27 38 BY SIMILARITY.
FT	DISULFID 32 55 BY SIMILARITY.
FT	DISULFID 57 66 BY SIMILARITY.
FT	DISULFID 70 225 BY SIMILARITY.
FT	DISULFID 212 216 BY SIMILARITY.
FT	DISULFID 230 387 BY SIMILARITY.
FT	CARBOHYD 238 238 POTENTIAL.
FT	CARBOHYD 325 325 POTENTIAL.
FT	CARBOHYD 329 329 POTENTIAL.
FT	CARBOHYD 350 350 POTENTIAL.
SEQ	SEQUENCE 387 AA; 43123 MW; 9672347E CRC32;

[illegible]

Db 215 ACTLRFELLGCELNGCANPLGLKNNISIPDKOITASSSYKTWGLHLF-SWNSYARLDKOG 273

QY 1340 RSNAMRPQVANNPKEMLOVDOKTMKVTGYTQGVKSLTSMYVKEFLISSQDGHOWTLF 1399

Db 274 NFNAMVAGSIGNDQWLQVDLGSSKEVTGITQGARNFGSVQFVASYKVAYSNDSANWTEY 333

QY 1400 F--QNGKVVYFQGNODSFTPVVNSLDPPLRLRYLRHPOSVWVHQIALRMEVLC 1451

Db 334 QDPRTGSSKIFPGNMWMDNHSKKNLFETPIARVIRLPVAMNRRIALRLLELGC 387

RESULT 14

NRP_CHICK ID NRP_CHICK STANDARD; PRT; 914 AA.

AC P9795;

DT 01-NOV-1997 (REL. 35, CREATED)

RT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

CC 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

NEUROFILIN PRECURSOR (A5 PROTEIN).

GN NRP.

OS GALLUS GALLUS (CHICKEN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;

OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-WHITE LEGHORN; TISSUE-EMBRYONIC BRAIN;

RX MEDLINE; 95324761.

RA TAKAGI S., KASUYA Y., SHIMIZU M., MATSUURA T., TSUBOI M., KAWAKAMI A.,

RT "Expression of a cell adhesion molecule, neuropilin, in the

RL DEV. BIOL. 170:207-222(1995).

CC -I- FUNCTION: CALCIUM-INDEPENDENT CELL ADHESION MOLECULE THAT FUNCTION

CC DURING THE FORMATION OF CERTAIN NEURONAL CIRCUITS.

CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC (LAYERS D AND E OF SGFS), AMACRINE CELLS OF RETINA, NEURITES OF

CC DORSAL ROOT GANGLIA. ALSO EXPRESSED IN NONNEURONAL CELLS, E.G.

CC BLOOD VESSELS IN THE ENTIRE EMBRYO.

CC -I- SIMILARITY: CONTAINS 2 CUB DOMAINS.

CC -I- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.

CC -I- SIMILARITY: CONTAINS 1 MAM DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; D45416; G1841311; -.

DR PROSITE; PS01180; CUB; 2.

DR PROSITE; PS01285; FA58C_1; 2.

DR PROSITE; PS01286; FA58C_2; 2.

DR PROSITE; PS00740; MAM_1; 1.

DR PROSITE; PS0060; MAM_2; 1.

DR PFAM; PF00431; CUB; 2.

DR PFAM; PF00629; MAM; 1.

DR PFAM; PF00754; F5_F8_type_C; 2.

KW TRANSMEMBRANE; GLYCOPROTEIN; NEURONE; SIGNAL; REPEAT; CELL ADHESION.

FT SIGNAL 1 18

FT CHAIN 19 914

FT DOMAIN 20 847

FT TRANSMEM 848 870

FT DOMAIN 871 914

FT DOMAIN 25 139

FT DOMAIN 145 263

FT DOMAIN 273 422

FT DOMAIN 429 581

FT DOMAIN 636 801

FT DISULFID 25 52

FT DISULFID 80 102

DR PROBABLE.

FT DISULFID 145 171 PROBABLE.

FT DISULFID 204 226 PROBABLE.

FT DISULFID 273 422 BY SIMILARITY.

FT DISULFID 429 581 BY SIMILARITY.

SQ SEQUENCE 914 AA; 102480 MW; 593B7139 CRC32;

Query Match 5.9%; Score 458.5; DB 1; Length 914;

Best Local Similarity 35.0%; Pred. No. 5.4e-22;

Matches 113; Conservative 58; Mismatches 129; Indels 23; Gaps 13;

QY 1145 KCQTPIGMASGHIRDFQITASGOYGO-WAPKLARLHSGSINAMSTKEPF--SWIKVDLL 1201

Db 272 QCMPELGESEIHSQITVSSQYSAIWSSERSRLNYPE--NGWTEGEDSVREWIQVDLG 329

QY 1202 APMIHGIKTQGA--RKFSSLYISQFIIMYSLDGKKWQTYR-GNSTGTLMVFFGVNDSS 1258

Db 330 LLRFVSGIGTQGAISKETKKEYTLKTYRVDVSSNGEDWITLKEGNKP--VVFQGNAPT 386

QY 1259 GIKHNIFNPPIARIYIRLPHYISIRSTLRMELMGDLNS--CSMPIGESKAISDAQIT 1316

Db 387 DVYRPPPKPVLTRFVRIRKPVSWENGSLRFVEYGEKITDYPSCGMIGVSGLIPDSQIT 446

QY 1317 ASSYFTNMFATWSPSKARLHLQGRSNAMRPQVNNP--KEWLQVDFOKTMKVTGYTQGVK 1374

Db 447 AS--TQVDRNMWIPENARL-ITRSRGVALPPTTHPTNENWLQIDLGEEKIVRGIIYQGGK 502

QY 1375 SLTSMYVKEFLISSQDGHOWTLFQNG--KKVFOGNODSFTPVVNSLDPPLRLRYLR 1432

Db 503 HRENKVFMRKFIKIGYNSNGSDWKIMDSKKIKITFGNTNTYDPELRTEE-PVSTRIR 561

QY 1433 IHPOSVWVH-QIALRMEVLCQEAQ 1454

Db 562 VYPERATHAGLRLMELGCELE 584

RESULT 15

NRP_MOUSE ID NRP_MOUSE STANDARD; PRT; 923 AA.

AC P97333;

DT 01-NOV-1997 (REL. 35, CREATED)

RT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

CC 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE NEUROFILIN PRECURSOR (A5 PROTEIN).

GN NRP.

OS MUS MUSCULUS (MOUSE).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BALB/C; TISSUE-EMBRYONIC BRAIN;

RX MEDLINE; 96353149.

RA KAWAKAMI A., KITSUKAWA T., TAKAGI S., FUJISAWA H.,

RT "Developmentally regulated expression of a cell surface protein,

RL J. NEUROBIOL. 29:1-17(1996).

CC -I- FUNCTION: CALCIUM-INDEPENDENT CELL ADHESION MOLECULE THAT FUNCTION

CC DURING THE FORMATION OF CERTAIN NEURONAL CIRCUITS (BY SIMILARITY).

CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -I- TISSUE SPECIFICITY: NERVOUS SYSTEM.

CC -I- SIMILARITY: CONTAINS 2 CUB DOMAINS.

CC -I- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.

CC -I- SIMILARITY: CONTAINS 1 MAM DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; D50086; E293356; -.

DR	MGD; MGI:106206; NRP.
DR	PROSITE; PS01180; CUB; 2.
DR	PROSITE; PS01285; FA58C_1; 2.
DR	PROSITE; PS01286; FA58C_2; 2.
DR	PROSITE; PS00740; MAM_1; 1.
DR	PROSITE; PS50060; MAM_2; 1.
DR	PEAM; PF00431; CUB; 2.
DR	PEAM; PF00629; MAM; 1.
DR	PEAM; PF00754; F5_F8_type_C; 2.
KW	TRANSMEMBRANE; GLYCOPROTEIN; NEURONE; SIGNAL; REPEAT; CELL ADHESION.
FT	SIGNAL 1 20
FT	CHAIN 21 923
FT	DOMAIN 21 856
FT	TRANSMEM 857 879
FT	DOMAIN 880 923
FT	DOMAIN 27 141
FT	DOMAIN 275 265
FT	DOMAIN 431 583
FT	DOMAIN 645 811
FT	DISULFID 27 54
FT	DISULFID 82 104
FT	DISULFID 147 173
FT	DISULFID 206 228
FT	DISULFID 275 424
FT	DISULFID 431 583
FT	CARBOHYD 150 150
FT	CARBOHYD 261 261
FT	CARBOHYD 300 300
FT	CARBOHYD 522 522
FT	CARBOHYD 842 842
SO	SEQUENCE 923 AA; 103019 MW; 265CODE7 CRC32;

Query Match	5.78;	Score 446.5;	DB 1;	Length 923;
Best Local Similarity	34.78;	Pred. No. 3.2e-21;		
Matches 112; Conservative	57;	Mismatches 131;	Indels 23;	Gaps 13;

```

QY 1145 KQOTPLGMA$GHRDFOITASQYQ-QWAPKLARLHYSGINAW$TKPEF--SWIKVDLL 1201
    11 111 11 11111 111 : : :11:1 11 : 1 11:111
Db 274 KCM$ALGME$GEIHS$DQITASSQYGTNMSVER$RNLNPE--NGWTPGED$YKEMIQVDLG 331
QY 1202 APMIIGHIKTQGA--RQKFSSLYISOFIIMYSLDGKKMOTYR-GNSTGTLMVFFGNVDS$ 1258
    : : 1111 : : : : : : 1 : 1 : : 11 : : 11 : :
Db 332 LIRFVTAVGTQGAISKETKKKYVKT YRVDISSNGEDWISLKEGNKA--11FQGTNPT 388
QY 1259 GIKHNIENPRIARYIRLHPTHYSIR$TLRMEIMGCDLNS--CSMPLGME$KAISDAQIT 1316
    : :1: 1: 1: :1: 1: : : :1: 1: 11 111 111 111
Db 389 DVLGVFSKPLITRFVRIRIKP$W$METGISMRFEYQCKITDYP$CSGMLGMY$GLISDSQIT 448
    1317 ASSYFTNMFATW$PSKARLHLQGR$N$N$M$R$Q$V$N$N$P--KEWLOVDQKTMKYTGVTQGVK 1374
    11: 11 11 11 : : :1 111111 11 11 11 11
Db 449 AS$NQADR--N$N$M$PENIRL-VISRTG$W$ALLP$P$D$P$H$PTN$EWLOVDLGDEKIRGVIIQGGK 504
    1375 SILTSMYKKEPLISSQDGHQWTLTFQNG--KYKVFQGNQDSETPVYN$LDPEPLITRYLR 1432
    : : : : 1: 1: 1: 1: : 1 11:11: 11 : : 11 11:1
Db 505 HREK$KVFMRK$F$K$AY$SNSNG$DM$K$T$IMD$D$K$R$K$A$K$F$EG$N$N$N$IDTPELRTFS-PL$TR$FIR 563
QY 1433 IHPQ$SWVHQ-IALRMEVLGCEAQ 1454
    1:1: 1 : 111:111:
Db 564 IYERATH$G$LG$LRM$ELLGCEVE 586

```

Search completed: August 19, 1999, 12:39:19
Job time: 2333 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 1999, 12:01:45 ; Search time 19.9 Seconds

(without alignments)
4505.984 Million cell updates/sec

Title: US-09-001-039a-47

Perfect score: 7797

Sequence: 1 MQIELSTCFFLCLRLFCFSA.....WVHQIALRMEVLGCEADLY 1457

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database :

- SPTREMBL_10:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6347	81.4	2343	6	018806	018806 canis famill
2	6306	80.9	2343	6	062730	062730 canis famill
3	2387.5	30.6	2224	4	043737	043737 homo sapien
4	2374.5	30.5	2183	11	088783	088783 mus musculu
5	1808	23.2	1157	11	092024	092024 mus musculu
6	1341	17.2	891	4	075180	075180 homo sapien
7	1230	15.8	782	4	075659	075659 homo sapien
8	1104	14.2	216	4	014286	014286 homo sapien
9	662.5	8.5	480	4	043854	043854 homo sapien
10	655.5	8.4	480	11	035474	035474 mus musculu
11	514.5	6.6	363	6	077718	077718 equus cabal
12	469.5	6.0	931	4	060462	060462 homo sapien
13	469.5	6.0	909	4	014820	014820 homo sapien
14	469.5	6.0	926	4	014821	014821 homo sapien
15	464.5	6.0	925	11	035276	035276 ratius norv
16	462.5	5.9	909	11	035373	035373 mus musculu
17	462.5	5.9	926	11	035374	035374 mus musculu
18	462.5	5.9	931	11	035375	035375 mus musculu
19	462.5	5.9	901	11	035376	035376 mus musculu
20	462.5	5.9	906	11	035377	035377 mus musculu
21	462.5	5.9	914	11	035378	035378 mus musculu
22	429.5	5.5	923	4	060461	060461 homo sapien
23	429.5	5.5	923	4	014786	014786 homo sapien
24	408.5	5.2	558	5	076470	076470 lytechinus
25	271	3.5	1128	11	088442	088442 mus musculu
26	266	3.4	764	11	054860	054860 mus musculu
27	262	3.4	845	4	014113	014113 homo sapien
28	247	3.2	224	11	092114	092114 mus musculu
29	243	3.1	224	4	015537	015537 homo sapien

30	223	2.9	728	6	097567	097567 bos taurus
31	223	2.9	722	11	092100	092100 mus musculu
32	221	2.8	719	11	061281	061281 mus musculu
33	211	2.7	1283	5	094887	094887 drosophila
34	207.5	2.7	737	5	017305	017305 caenorhabdi
35	206.5	2.6	737	5	018163	018163 caenorhabdi
36	185.5	2.4	911	11	035407	035407 mus musculu
37	183	2.3	1384	4	078357	078357 mus musculu
38	171	2.2	1385	11	054991	054991 mus musculu
39	171	2.2	1381	11	097846	097846 ratius norv
40	156.5	2.0	78	4	092247	092247 homo sapien
41	153.5	2.0	574	10	092247	092247 homo sapien
42	143.5	1.8	1196	9	092247	092247 clostridium
43	141.5	1.8	1196	2	045916	045916 clostridium
44	141.5	1.8	1196	9	038197	038197 clostridium
45	140.5	1.8	1217	2	045440	045440 bacillus sp

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	2343 AA.
ID	018806			
AC	018806:			
DT	01-JAN-1998 (Tremblrel. 05, Created)			
DT	01-JAN-1998 (Tremblrel. 05, Last sequence update)			
DT	01-MAY-1999 (Tremblrel. 10, Last annotation update)			
DE	FACTOR VIII.			
GN	F8.			
OS	Canis familiaris (Dog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER;			
RA	CAMERON C., NOTLEY C., HOYLE S., MCGLYNN L., HOUGH C., KAMISUE S.,			
RA	GILES A., LILLICRAP D.;			
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF016234; AAB87412.1; -			
DR	PFAM; PF00394; Cu-oxidase; 3.			
DR	PFAM; PF00754; F5_F8_type_C; 2.			
DR	PROSITE; PS01285; FA58C_1; 2.			
DR	PROSITE; PS01286; FA58C_2; 2.			
SQ	SEQUENCE 2343 AA; 265829 MW; 97COAF09 CRC32;			

Query Match 81.4%; Score 6347; DB 6; Length 2343;
Best Local Similarity 53.6%; Pred. No. 0;
Matches 1259; Conservative 83; Mismatches 105; Indels 900; Gaps 4;

QY	1	MQIELSTCFFLCLRLFCFSATRRYYIGAVELSDYMQSD-LGELPVDARFPPRPKSPF	59
DB	1	MQVELYTCFCFLCLLPFLSATRRYYIGAVELSDYMQSDLSALHADTSFSSRPGLPL	60
QY	60	NTSVYKKTLFEVETDHLFNIAKPRPPMGLGPTIQAEVYDTVYITLKNMASHPVSLHA	119
DB	61	TTSVTYRKTYFEFTDLDLNFNAKPRPPMGLGPTIQAEVYDTVYIVLKNMASHPVSLHA	120
QY	120	VGVSYWKASEGAEYDDQTSQREKEDDKVPPGSHYTYWQVLKENGPMASDPCLTYSYLS	179
DB	121	VGVSYWKASEGAEYEDQTSQREKEDDNVIPGESHYTYWQVLKENGPMASDPCLTYSYFS	180
QY	180	HYDLVKDLNSGLIGALLVCRBSLAKETQTLHKFILLFAVDEGKSWHSETKNSLMQDR	239
DB	181	HYDLVKDLNSGLIGALLVCKEGLAKERTQTLQEFVLLFAVDEGKSWHSETNASLTO--	239
QY	240	DAASARAWPKMTVNGYVNRSLPGLIGCHRSYVWHVIGMGTTPVHSHIFLEGHTFLVRN	299
DB	239	---AEOHELTINGYVNRSLPGLITVCHKRSYVWHVIGMGTTPVHSHIFLEGHTFLVRN	294
QY	300	HROASLEISPTFLTAQTLLMDLGQFLFCHTSSHQHDMGEVYKVDSCPEPQLRMKN	359

Db	295	HRQASLEISPTFFLLTAQTFLMDLGOELLFCHIPSHQHDGMEAYVKVDSCPEEPOLRMKN	354
QY	360	EEAEDYDDDLTDSEMDVYRFDDDNSPSTQIRSVAKKHPTWVHYIAAEEDWDYAPLV	419
Db	355	ED-KDYDDGLYSDMDVVSFDDSSPFIQIRSVAKKHPTWVHYIAAEEDWDYAPSGP	413
QY	420	APDDRYSKQYLNNGPQIRGRYKKVRFMATYDTEFTKTRAIQIHESGILGPLLYGEVGD	479
Db	414	TPDRSHKNLYLNGPQIRIGKTKKVRFAVAYTDETFKTRAIQYESGILGPLLYGEVGD	473
QY	480	LLIFKNOASRPNIYPHGITDYRPLYSRRLPKGVKHLKDFILPGEIFKWTVTVEG	539
Db	474	LLIFKNOASRPNIYPHGINVYTPRLHTGRLPKGVKHLKDMFILPGEIFKWTVTVEG	533
QY	540	PTKSDPRCLTRYSSFVNMERDLASGLIGPLLICYSVDQROGNQIMSDKRNVLFSVD	599
Db	534	PTKSDPRCLTRYSSFINLERDLASGLIGPLLICYSVDQROGNQIMSDKRNVLFSVD	593
QY	600	ENRSMYLTENIQRELPNPAGVQLEDEPEFQASNMHSINGYFDSLQLSVCLHEVAYWIL	659
Db	594	ENRSMYLTENMQRELPNADVQHPDEPQLSNIMHSINGYFDMQLSVCLHEVAYWIL	653
QY	660	SIGAOTDFLSVFFSGYTFKHKNVYEDTLTLPFSGETVFMSENPGWLWICHSNDFRNR	719
Db	654	SVGAOTDFLSVFFSGYTFKHKNVYEDTLTLPFSGETVFMSENPGWLWVGCHNSDFRNR	713
QY	720	GMTALLKVSSCDKNTGDYEDYESYEDISAYLLSKNNAIEPRSF-----	762
Db	714	GMTALLKVSSCNRNIDYEDYEDIPRTLNNVNIKPRSFQNSRHPSTREKQKAT	773
QY	762	-----	762
Db	774	TPENDIEKIDQSGERTOLIKAQSVSSDLMLLGONPTPRGLFLSDLRATDRADHSR	833
QY	762	-----	762
Db	834	GATERNKGPREVASLRLRHSEDREFTPPELOLRNLNENLGINTVELKLDKISSSS	893
QY	762	-----	762
Db	894	DSLMTSPTIPSDKLAATEKTSGLCPPNMSVHNSHLGITVEGNSSSHLQSGVPLELSE	953
QY	762	-----	762
Db	954	EDNDSKLLIAPLMIQOESSLRENVLSMESNRLEKFEERIRGPASLIKDNALFKVNISVKT	1013
QY	762	-----	762
Db	1014	NRAPVNLTTNRKTRVAIPTLLIENSTSWQDIMLERNTEFEKVTSLIHNETFMDRNTAL	1073
QY	762	-----	762
Db	1074	GLNHSNKTTLSKNVEMAHQKEDBPVPLRAENPDLSSKIPFLPDWIKTHGKNSLSSEQR	1133
QY	762	-----	762
Db	1134	PSPKQLTSLGSEKSVKQDNFLSEKVVGEDEFTKDTLQEIIFPNKKSIFANLANVOEN	1193
QY	762	-----	762
Db	1194	DTYNOEKSPPEIERKEKLTQENVALPOAHTMIGTKNFLKNLFLSTKQNVAGLEEOPYT	1253
QY	762	-----	762
Db	1254	PILQDTRSLNDSPHSEGIHMANFSKIREANLEGLGNQTNQOVERPSTTRMSSNASQHV	1313
QY	762	-----	762
Db	1314	ITQGRKSLKQPRLSQGEIKFERKVIANDTSTQWKNMNYLAQGTLTQIEYNEKEKRAIT	1373
QY	762	-----	762
Db	1374	QSPLSDCSMRNHVTIQMNDSALPVAKESASPSVRHTDLTKIPSOHNSSHLPASACNYTFR	1433

QY	762	-----	762
Db	1434	ERTSGVQEGSHFLQEAKRNNLSLAFVTLGITTEGQGFSSJGKSATNOPMAYKLENTVLLQ	1493
QY	762	-----	762
Db	1494	PGLSETSDKVELLSQVHVQDEDSFPTKTSNDSPGHLDMGKIFLQKTQGPVKMKNKTNSPG	1553
QY	762	-----	762
Db	1554	KVPELKWATESSEKIPSKLLGVLAWDNHDTQIPSEEWKSQKKSQTNTAFKRKDTILPLG	1613
QY	762	-----SÖNPVLKRHÖREITRTTLQSDQEEI	787
Db	1614	PCENNDSIAINEGQDKPQREAMAKQGEPRGLCSQNPVYSKHHÖREITVTLQPEEDKF	1673
QY	788	DYDDTISVEMKKEDEPDYDEDENOSPSPFOKTRHYFIAAVERLMDYGMSSPHVLNRRA	847
Db	1674	EYDDTFSEMKREDEPDYGYDENOGLRSFOKTRHYFIAAVERLMDYGMSSPHILNRRA	1733
QY	848	QSGSPQPFKKVVPQEFDTGSEFTQPLRYGELNEHLLGPRYTRAVEDNIMVTERNOASRP	907
Db	1734	QSGDVQQFKKVVQEFETDGSFTQPLRYGELNEHLLGPRYTRAVEDNIVTFKNQASRP	1793
QY	908	YSFYSLLISYEEDQROGAEPKRKNPVKPNETKTYFWKVQHMAPTKDEDFCKAMAYFSDDV	967
Db	1794	YSFYSLLISYDEDEGQGAEPKRKRVNPNETKIYFWKVQHMAPTKDEDFCKAMAYFSDDV	1853
QY	968	LEKDVHSGLLGPLLYCHTNTLNPAHGRQVTVQEFALFTTIPDETYSWYFTENMERNCRAP	1027
Db	1854	LEKDVHSGLLGPLLICRSNTLNPAHGRQVTVQEFALVFTTIPDETYSWYFTENLERNCRAP	1913
QY	1028	CNIOWEDPTFEKENRFHAINGYIMDTLPGLYMAQDORIRWYLLSMGSNENIHSHIFSGHV	1087
Db	1914	CNVQKEDPTLKENRFHAINGYKDTLPGLYMAQDQKVWYLLSMGSNENIHSHIFSGHV	1973
QY	1088	FTVRKKEEYKMATLYNLYPGVFETVEMLPSKAGIMRVECLIGEHLHAGMSTLEFLVYSNKCQ	1147
Db	1974	FTVRKKEEYKMAVYNLYPGVFEVEMLPSQVGIMRIECLIGEHLQAGMSTLEFLVYSKCKQ	2033
QY	1148	TPLGMA SGHIRDFQITASGOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIH	1207
Db	2034	TPLGMA SGHIRDFQITASGOYGOWAPKLARLHYSGSINAWSTKDPFSWIKVDLLAPMIH	2093
QY	1208	GIMTOGARQKFSLSLYISOIIMYSLIDGKKWQTYRGNSTGTLMVFFGNVDSGCIKHNIENP	1267
Db	2094	GIMTOGARQKFSLSLYSQOIIIMYSLDGNKMWHSYRGNSTGTLMVFFGNVDSGCIKHNIENP	2153
QY	1268	PIIARYIRLHPHTHYSIRSTLRMELMGCDLNSCSPDLGME SKAISDAQITASSYFTNMFAT	1327
Db	2154	PIIAQYIRLHPHTHYSIRSTLRMELMGCDENSCSPDLGME SKAISDAQITASSYLSM LAT	2213
QY	1328	WSPSKARLHLOGRSNAMPPOVNNPKFWLQVDFQKTMKVTVGTTQGVKSLTSMYKFEFLI	1387
Db	2214	WSPSQARLHLQGRITNAMPPOANNPKFWLQVDFRKTMKVTGITQGVKSLTSMYKFEFLI	2273
QY	1388	SSSQDGHQWTLFFQNGKVKVVFQGNODSETPVVNSLDPPLLTRYLRIHPQSWYHQIALRME	1447
Db	2274	SSSQDGHNWTLEFLQNGKVKVVFQGNODSSTPVVRNLEPPLVARIYVRLHPQSWAHHIALRLE	2333
QY	1448	VLCGEAQ 1454	
Db	2334	VLCGDTQ 2340	

RESULT	2	
062730		
ID	062730	PRELIMINARY;
AC	062730;	PRT; 2343 AA.
DT	01-AUG-1998	(TREMBLrel. 07, Created)
DT	01-AUG-1998	(TREMBLrel. 07, Last sequence update)
DT	01-MAY-1999	(TREMBLrel. 10, Last annotation update)

DE FACTOR VIII.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=KIDNEY, SPLEEN;
RC GORDY P.W., BOWEN R.A.;
RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF049489; AAC05384.1; -.
DR PFAM; PF00394; Cu-oxidase; 3.
DR PFAM; PF00754; F5_F8_type_C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
SQ SEQUENCE 2343 AA; 265613 MW; D9C5A549 CRC32;

Query Match	80.9%;	Score 6306;	DB 6;	Length 2343;
Best Local Similarity	53.3%;	Pred. No. 0;		
Matches 1252;	Conservative	86;	Mismatches 109;	Indels 900;
				Gaps 4;

1	MOUELTCEFLCLLRECFESATRRYYIGAVELSWDYMQSD-LGELPYDAREPPRVPKSPF	59
Db	1 MOUELYTCCEFLCLLPFSLSATRKYYIGAVELSWDYMQSDLLSALHADTSFSSRVPGSLPL	60
QY	60 NTSVYVKKTLFVEFTDHLFNIAKPRPBMGLPPTIOAEVYDTVITLKNMASHPVSLHA	119
Db	61 TTSVYTRKTVFVEFTDLEFNIAKPRPBMGLPPTIOAEVYDTVITLKNMASHPVSLHA	120
QY	120 VGVSTWKAASEGAEDDQTSOREKEDDKVFPGGSHTYVQVLKENGPMASDPLCLTYSLS	179
Db	121 VGVSTWKAASEGAEDDQTSOREKEDDNIIPGESHYVQVLKENGPMASDPLCLTYSFS	180
QY	180 HVLDYKDLNSGLIGALLVCREGLAKEKTQTLHKFILLFAVFEDEGKSMHSETKNSLMQDR	239
Db	181 HVLDYKDLNSGLIGALLVCKEGLAKERTQTLQEFVILLFAVFEDEGKSMHSETNASLTQ--	239
QY	240 DAASRAMPKMHTVGVNRSPLPGLIGCHRSVYWHVIGMGTPEVHSIFLEGHTFLVGN	299
Db	239 ---AEAQHELHTINGVYVNRSLPGLYVCHKRSVYWHVIGMGTPEVHSIFLEGHTFLVGN	294
QY	300 HRQASLEISPTIFLTAOTFLMDLGQFLFCHISSHQDGMAYKVVDSCPEEPOLRMKN	359
Db	295 HRQASLEISPTIFLTAOTFLMDLGQFLFCHIPSHQDGMAYKVVDSCPEEPOLRMKN	354
QY	360 EEAEYDDDDLTDEMDVYRFDDDNBSFIQIRSVAKNHPKTWVHYIAAEEDMDYAPLV	419
Db	355 ED-KDYDDGLGSDMDVVSFDDSSSPFIQIRSVAKNHPKTWVHYIAAEEDMDYAPSGP	413
QY	420 APDBRSYKSOYLNNGPORIGRKYKKVFMAYTDETFKTREAIQHESGILGPLLYGEVDT	479
Db	414 TPNDSSHKNLYLNNGPORIGKYYKKVFAVYTDTEFKTREAIQYESGILGPLLYGEVDT	473
QY	480 LLITRKQASRPYNYPHGITDVRPLYSRRLPKGVKHLKDFILPGETFYKKWTYVEDG	539
Db	474 LLITRKQASRPYNYPHGINVYTPLHTGRLPKGVKHLKDMFILPGETFYKKWTYVEDG	533
QY	540 PTKSDPRCLTRYSSFEVNMERDLASGLIGPLLICYESVDQRGNOJMSDKRNVILFSVD	599
Db	534 PTKSDPRCLTRYSSFEINLERDLASGLIGPLLICYESVDQRGNOJMSDKRNVILFSYLD	593
QY	600 ENRSWYLTENIQRFLEPNPAGVQLEDPFEQASNIMHSINGVYFDSLQLSVCLHEVAYWIL	659
Db	594 ENRSWYLTEDMQRFLEPNADVQPHDPEQLSNIMHSINGVYFDTQLSVCLHEVAYWIL	653
QY	660 SIGAOTDFLSVFFSGYTFKKHMYEDTLTFPESGETVEMSMENPGLMILGCHNSDFNR	719
Db	654 SVGAQOTDFLSVFFSGYTFKKHMYEDTLTFPESGETVEMSMENPGLMVLGCHNSDFNR	713
QY	720 GMTALLKVSSCDKNTG DYEDSYEDISAYLLSKNNAIEBRSF-----	762
Db	714 GMTALLKVSSCNRIDYEDYEDYEDIPPLNENNVIPRSFSQNSRHSSTREKOLKATT	773

QY	762	-----	762
Db	774	TPENDIEKIDQSGERTOLIKAQSVSSDDLMLLGONPTPRGLFLSLDREATDRADHSR	833
QY	762	-----	762
Db	834	GAIERNKGPPEVASLRPELRHSEDRFTPEPEQLRLNENIGNTTVELKLDKISSSS	893
QY	762	-----	762
Db	894	DSLMTSPTIPSDKLAATEKTGSLGRPNMSVHFNHGLGTIVFGNNSSHLIQSGVPLELSE	953
QY	762	-----	762
Db	954	EDNDSKLLAPLMIHQESSLRENYLSMESNRLEKEERIRGPASLIKDNALFKVNISVKT	1013
QY	762	-----	762
Db	1014	NRAPVNLTTNRKTRVAIPTLLIENSTSVWQDIMLERNTEPEKVTSLIHNETFMDRNTAL	1073
QY	762	-----	762
Db	1074	GLNHVSNKTTLSKNVEMAHQKKEDPVPRLAENBDLSSKIDPELPDWIKTHGKNSLSSEQR	1133
QY	762	-----	762
Db	1134	PSPKQTLTSGSEKSVKQDNFLSEKVVVGEDFTKDTLEQEIFPNKKSIFEANLANVQEN	1193
QY	762	-----	762
Db	1194	DTYNQEKKSLEIEERKEKLTQENVALPQAHTMIGTKNFLKNLELLSTKQNVAGLEEQPYT	1253
QY	762	-----	762
Db	1254	PILQDTRSLNDSPHSEGIHMANESKIREANIEGLGNQTNQWERFPSTTRMSSNASQHV	1313
QY	762	-----	762
Db	1314	ITQGRKSLKQPRLSQGEIKFERKVIANDTSGQWSKNMNYLAQTLTQIEYNEKEKRAIT	1373
QY	762	-----	762
Db	1374	QSPLDSCSMRNHVTTIQMNDALPYAKESASPSVYRHTDITGKIPSQHNSSHLPASACNYTFR	1433
QY	762	-----	762
Db	1434	ERTSGVQEGSHFLQEAKRNNLSLAFVTLGITGEGQGFSSLSGKSATNQPMYKKLENTVLLQ	1493
QY	762	-----	762
Db	1494	PGLSETSDKVELLSQVHVDQEDSEFPKTSNDSPGHLDMGKIFLQKTQGPYKMNKNTNSPG	1553
QY	762	-----	762
Db	1554	KVPFLKWATESSESEKIPSKLLGLVLAWDNHYDTQIPSEEMKSQKKSQTNIAFKRKDITLPLG	1613
QY	762	-----SONPVLKRHQREITRTTLQSDQEEI	787
Db	1614	PCENNDSTAINEGQDKPQREAMMAKQGEPRGLCSQNPVSKNHQREITVTLQPEEDKF	1673
QY	788	DYDDTISVEMKKEDFDIYDEDENOSPRSFOKKTRHYFIAAVERLMDYGMSSSPHVLNRRA	847
Db	1674	EYDDTFESIEMKREDFDIYGDYEDQGLRSFQKKTRHYFIAAVERLMDYGMSSSPHILNRRA	1733
QY	848	QSGSVQPKKVVFOEFTDGSFTQPLYRGELNEHLGLGPYIRAEVEDNINWTFRNQASRP	907
Db	1734	QSGDVQOEFKKVFOEFTDGSFTQPLYRGELNEHLGLGPYIRAEVEDNIVTFEKNQASRP	1793
QY	908	YSFYSSLSIYEEDQOGAEPKRKNFVKPNETKTYFWKVQHMAAPT KDEFDCAWAYFSDDVD	967
Db	1794	YSFYSSLSIYDEDEGOGAEPKRKRFPVNPNETKIYFWKVQHMAAPT KDEFDCAWAYFSDDVD	1853
QY	968	LEKDVHSGLIGPLLVCHTNTLNPAGHQVTVQEFALFFTIFDETKSWYFTENMERNCRAP	1027

Db 1163 HKSFTDISQSPSSHEWQTVISPDLSQVTLSPELSQTNLSPDISHTTLPSELIQRL 1222
QY 772 ----- 772
Db 1223 SPALGOMPISPDLSHTTLPDLSTHTTSLDLSQTNLSPELSQTNLSPALGOMPLSPDL 1282
QY 772 ----- 772
Db 1283 TTLSLDFSQTNLSPELSHMTLSPELSQTNLSPALGOMPISPDLSHTTSLDFSQTNLSP 1342
QY 772 ----- 772
Db 1343 LSQTNLSPALGOMPLSPDPSTHTTSLDLSQTNLSPELSQTNLSPDLSEMPLEFADLSQIPL 1402
QY 788 ----- 788
Db 1403 TPDLQMTLSPDLGETLSPNFGQMSLSPDLQVTLSPDISDTTLPLDLSQLSPDDLQ 1462
QY 788 ----- 788
1463 IFPSESSQSLLLOEFENESFPYPLDQOMPSPTLNDTFLSKENPLVIGLSKGTDY 1522
793 ISVENKKE-----DFDIYDE-----DENQSP-----RSFOKTRHY 823
1523 IEIIPKEEVOSSEDDYAEIDYVPYDDPYKTDVRTNINSSRDPDNIAWYLRNNGNRNY 1582
QY 824 FIAAVERLMDYGMSSSPHVLNRN--AQSGSVPO--FKVVFQEFDTGDSFTQPLRNGELN 878
1583 YIAAEIISWDY-----SEFQRETIDSDDDIPEDTYYKKVFRKYLDTSTFKRDPREYE 1638
QY 879 EHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLSYE-----EDQROGAEDPRKN 930
Db 1639 EHLGLGPYIRAEVDYVQVRFKNLARSPLHAHGLSYEKSSEKTYEDDSPEWERDN 1698
QY 931 FVKPNETKTYFWKVQHMAPTKDEFQCKAWAFSDVLEKDVHSGLIGPLLVCHTNTLNP 990
Db 1699 AVQPNSSYTYVWHATERSGPSPGACRAWAYSAVPEKDIHSGLIGPLLICQGLLHK 1758
QY 991 AHGROYVQEFALFTIETKSWYFTEEMERNCRAPCNIQMEDPTFKENYRFHAINCYI 1050
Db 1759 DSNMMDMREFVLFTMTFDEKKSWEKKSRSWR-----LTSSEMKKSHEFHAINCMI 1812
QY 1051 MDPLBGLVMAQDQIRIRWYLLSMGSENINHSIHFSGHVFTVRKKEEKYKALNYLPVGET 1110
Db 1813 Y-SLPGLKMYEQEWVRLHLNLIGGSODIHVVHFGQTLLENGNKQHLQGLVWPLLPGSFKT 1871
QY 1111 VEMLPKAGIWRVECLIGEHLHAGMSTLLFVYSNKCOTPLGMAHGIRDFOITASGOYQ 1170
Db 1872 LEMKASKPGWMLNTEYGENQIRAGMOTPLIMDRDCRMPMGLSTGITSQIKASEPLGY 1931
QY 1171 WAPKLARLHSGSINAWSTKE--PRS--WIKVDLAPMIHIGIKTOGAROKFSSLYIS 1224
Db 1932 WEPLRLANNGGSYNAWSVEKLAERASKPWIQVDMQKEVITITGLOTQGAKHLYLKSCYTT 1991
QY 1225 QFTIMYSLDGKKWQYRGNSTGTLMVFFGNVDSGCIKHNIFNPPIIARYIRLHPTHSIR 1284
Db 1992 EEFYAVSSNOIMQOIFKGNSTRNVMYFNGNSDASTIKENQFDPPIVARYIRISPTRAYNR 2051
QY 1285 STLRLMLGCDLNSCMLPGLMESKASIDAQITASSYFTNMFAT--WSPSKARLHLQGRSNA 1343
Db 2052 PTLRLLELQCEVNGCSTPLGMEKNGIKENKQITASSFKKSWMGDYWEPRARLNAQGRVNA 2111
QY 1344 WRPOVNNPKEMLOVDEOKTMKVTGTQGVKSLTSMYVKEFLISSQDGHQWTLFEONG 1403
Db 2112 WQAKANNKQWLEIDLKIKKITAITQGCCKSLSEMYVKSYYTIHYSQGVEMKPYRLKS 2171
QY 1404 KV-KVFOGNQDSFTPVVNSLDPLRLTRYLRIRHPOSVWHQIALRMEVLCCEAODLY 1457
Db 2172 SMVDKIFEGNTNTKGHVKNFNPPIISRIFRIVIPKTNQSIALLRLELFGC--DIY 2224

RESULT 4
088783

ID 088783 PRELIMINARY; PRT; 2183 AA.
AC 088783;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE MURINE COAGULATION FACTOR V.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA YANG T.L., CUI J., REHMTULLA A., MOUSSALLI M., KAUFMAN R.J.,
RA GINSBURG D.;
RT "The structure and function of murine factor V and its inactivation
RT by protein C.";
RL Blood 91:0-0(0012).
RN [2]
RP SEQUENCE FROM N.A.
RA GINSBURG D., YANG T.L., CUI J., YANG A.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U52925; AAC99553.1; -
DR PFAM; PF00394; Cu-oxidase; 3.
DR PFAM; PF00754; F5_F8_type_C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
SQ SEQUENCE 2183 AA; 247228 MW; 5F22DD6A CRC32;

Query Match 30.5%; Score 2374.5; DB 11; Length 2183;
Best Local Similarity 25.7%; Pred. No. 1.8e-171;
Matches 582; Conservative 291; Mismatches 491; Indels 899; Gaps 33;

QY 8 CFLLCLL--RFC-----FSATRRYYLGAVELSWDYMOSDLGELPYDARFPRVPKS 56
Db 7 CFFLLVLTGRWAGMGSQAELAQLRQFYVLAQGLIMNYHE-----PTD 51
QY 57 FPEPNTSVYKKTLEVEFTDHLFNIAKPRPPWMLGPTIQAEVYDTVITLKNASHPV 116
Db 52 PSLNSIPSFKKIYREY-EQYFKKEKPRSSNSGLGPTLYAEVGVYKIVHFRNKADKPLS 110
QY 117 LHAAGVSYKASEGAEYDDQTSQREKEDDKVFPGSSHTYVWQYLKENGPMASDPLCTYS 176
Db 111 IHPQGIKYSKSEGASADHTFPAERKKDDAVAPGEYTYEWIVSEDSGPTPDDPCLTTHI 170
QY 177 YLSHVDLVKDLNSGLIGALLVCREGSLAKEKQTL--HKFILLFAVDEGKSWHSETKNS 234
Db 171 YSYENLTQDFNSGLIGPLLCCKGTLTDEGTQKMFQKQHVLLFAVDESKS----- 223
QY 235 LMQDRDAASARAWPMKHTVNGYVNSLPLGLIGHRSVYWHVIGMGTTPPEVHSIFEGHT 294
Db 223 -----RSQSPSLMYTTINGFVNKTMPTDITYCAHDHVSWMHLIGSSGPFLFSIHNGQV 274
QY 295 FLVRNHRQASLEISPTFLTAQTLMDLGOFLLECHISSHQHDMGMAVYKVDSCPEEPQL 354
Db 275 LEQNHQKSVSTVTLVSATSTTANMTMSPEGRWIVSSLI PKHYQAQAYIDIKNCPRKTR- 334
QY 355 RMKNNEEAEDYDDDLTDSEMDVVRPDDDNSPSFIQIRSVAKKHPRKTWVHYIAAEEDWDY 414
Db 334 -----SPK--TLTREQRRYMKRWEYFIAAEVYIWNV 362
QY 415 APLVLAPODRSYKSOYLNGPQIRGRKKYKVRFMAYTDEF--KTRERAIQHESGILPPL 472
Db 363 APVIRPAMMDKIYRSQHLDNFSNQIGKHYKVIYRQYEEETFTKRDNPSIKQSGILPVI 422
QY 473 YGEVGDLLIIFKNQASRPYNIYPHIGITDVRPLYSRLLPKGV-----KHLKDPILPG 525
Db 423 RAQVRDTLKIIVFKNMAASRPYSIYPHGVY-----FSPYEDGINSSTSGSHTTIRVQPG 476
QY 526 EIFKYKMTVTVEDGPTKSDPRCLFRRYSSFVNMERDLASGLIGPLLCYKESVDQGNQI 585
Db 477 ETFTYKWNILEFDEPTENDAQCLFRPYSDVDVTRDIASGLIGLLICKRSRLDQGVQR 536
QY 586 MSCKRNVILFVFDENRSWYLTENIQRLPNPAGVQLEDPEFQASNIMHSINGYVFDLSQ 645

[illegible]

```

1 |||: | | | : | : | | | | | | | : | :
Db 1029 EVEVEVNASNPGTWLMHCHVTDDHVAHAGMETIFTVLSHE 1065

RESULT 6
075180,
ID 075180 PRELIMINARY; PRT; 891 AA.
AC 075180;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE KIAA0698 PROTEIN.
GN KIAA0698.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=BRAIN;
RC MEDLINE; 98403880.
RX ISHIKAWA K., NAGASE T., SUYAMA M., MIYAJIMA N., TANAKA A., KOTANI H.,
RA NOMURA N., OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
DR EMBL; AB014598; BAA31673.1; -.
DR PFAM; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
SQ SEQUENCE 891 AA; 100624 MW; 4EA30F16 CRC32;

```

```

Query Match                               17.2%; Score 1341; DB 4; Length 891;
Best Local Similarity   31.1%; Pred. NO. 1.9e-93;
Matches    321; Conservative 148; Mismatches 358; Indels 206; Gaps 23;

QY      250 MHTVNGVYNRSLPGLIGCHRSKYVMHVIGMTTPEVHSIFLEGHFLVRNHRQASLEISP 309
       || :||| : || | : || | : || | : || | : || | : || | : || | : || |
Db       1 MHAINGFVFGNLPFLNMCAQKRVAMHLFGMGNEIDVHTAFHGMLTTRGHHTDVANIFP 60

QY      310 ITFTLAQTLLMDLIGQLLFCHISSHQHDGMEAAYKVDSCEEPOLRMKNNEADYDDDL 369
       ||| :|| : : || : || : || : || : || : || : || : || : || : || : ||
Db       61 ATFTYAEMVPWEPTGWLLISCQVNSHFRDGMQALYKVKSCSMAPPY-----DLL 108

QY      370 TDSEMDVVRFDDBNSPSFIQIRSYAKKHPKTWVHYIAEEDWDYAPL-----VLA 420
       - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Db       109 TG-----KVROYFEAHEIQWGYGPMGHDSTCKNLRE 141

QY      421 PDRRSYKSQYLNNBPQRIGRKYYKRYFMATYTDETFKTREAIOHES--GILGPLLYGEVD 478
       - - - - - : : : ||| | ||| : ||| : : : ||| : ||| : ||| : |||
Db       142 PG--SISDKFFOKSSSRIGGTMYKRYEAFODETPOEKMHLEBRHLGILGPVIRAEVD 199

QY      479 TLLITFKNQASRPNIYPHGITDYVRPLYSRRLPKGVKHLKDPIPLGELFK-----YKW 532
       || : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db       200 TIQVIFYNRASQPSFMOPHGV-----FYEKDYEGIV--YNDGSSTPGLVAKPFKEVTYRW 252

QY      533 TVTVEDGPTKSDPRCLTRYSSFFVMNERDLASGLIGPLLIYKSVYDQRGNOIMSDKRN 592
       || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       253 TVPPHAGPTAQDPACLTWMYFSAADPIRDNTSGLVPLVCBAGALGADGKKQGVDEKFF 312

QY      593 ILFSVFEDENRSWYLETENIQRFLPPAGVOL-----EDPE-FQASNIMHSINGVYFDSL- 645
       :||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db       313 LLFTVLDENKSWYSN-----ANOAAAMLDFRLISEDIEGFODSNRHAINGFLFSNLP 365

QY      645 QLSVCLHEVAIYWYIISIGAOTDLFSVFPFSGYTERHKMVEYEDTLTFPESGETVFMSEMNP 704
       :|| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db       366 RLDMCKGDTVAMHLLGLGETETDVHGVMFQGNIVOLQCMRKGAAMLFPHTFVMAIMQPDNL 425

QY      705 GLMITGCNDSDFRNRGMTALLKYSSCDKNTGDYIEDSYEDISAVLLSKNNAIEPRSFSQN 764
       | : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db       426 GTFEITYCQAGSHREGRAIYNSQC----- 452

```

QY 765 PVLKRHOREITRTTLQSDQEIDYDDTISVEMKKEDEDIYDEDENQSPRSFOKTRHYE 824
Db 452 -----PGHQ-----ATPRORYQAAARIYY 469
QY 825 IAAVERLMDYGMSSS-PHYLRNRAQSGSV-----POFKKVFQOFTDGSFTQ 870
Db 470 IMAEEVEMDYCPDRSWEREMHNSQEKDSYGYIFLSNKDGLGSRYKRAVREYTDGTFRI 529
QY 871 PLYRGELNEHLGLGPYIRAEVEDNINVTFRNQAASRPYSFYS-SLISYEEDQROGAEPK 929
Db 530 PRPRGPEEHLGLGPLIKGEVDILTVEFKNNASRPYSVHAHGVLESTVWPLAE--- 587
QY 930 NEVKPNETKTYFWKVQHMAPTKDEDFCKAWAYFSDVDLEKDVHSGILGPLLVCHNTLN 989
Db 587 ---PGEVVTYQWNIPERSGPGPNDSACVSWIYSAVDPIKDMYSGLVGLAICQKILE 642
QY 990 PAHGRQYTVQEFALFETFEDETSWYTEMN-ERNCRAPCNIQMEDPTFKENYRFAHNG 1048
Db 643 PHGRSDMDREFALLFLIFDENKSWYLEENAVATHGSQDPGSINLODETFLESNMKHAING 702
QY 1049 YIMDTLPGLVMAQDORIRMYLLSMGSNENIHSIHFSGHVFTVRKKEEKMALYNLYPGVF 1108
Db 703 KLYANLRGLTMYQGERVAMWMLAMQDVLHTIHFHASEFLYLRNGENTRADVDLFPPTF 762
QY 1109 ETVEMLPKAGIWRVECLIGEHLHAGMSTLLFVYSNKCQTPGLMASGHTRDFQITASGOY 1168
Db 763 EVEMVASNPGTWMHCHVTDVHAGMETLFTVFSR-----TEHLSPLTVITKETE 813
QY 1169 GQWAPKLARLHYSGSINAMSTKEPFSWIKVDLLAPMIHGIKTQGARQKFSLSYISQFI 1228
Db 814 KAVPPRDIE---EGNVKMLGMQIPK--NVELASVLV-----AISVTLILV 855
QY 1229 MYSLDGKKWQTYR 1241
Db 856 VLALGGVWYQHR 868

RESULT 7
ID 075659 PRELIMINARY; PRT; 782 AA.
AC 075659;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, last annotation update)
DE DJ46618.1 (COAGULATION FACTOR V
(ACTIVATED PROTEIN C COFACTOR), COAGULATION FACTOR VIII
(PROCOAGULANT COMPONENT) AND CERULOPLASMIN
(EC 1.16.3.1, FERROXIDASE) LIKE) (FRAGMENT)).
GN DJ46618.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA WILKINSON J.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL030998; CAA19742.1; .
DR PFAM; PF00394; Cu-oxidase; 2.
FT NON_TER 1
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 782 AA; 87939 MW; 7DD1IE52 CRC32;

Query Match 15.8%; Score 1230; DB 4; Length 782;
Best Local Similarity 32.8%; Pred. No. 4.3e-85;
Matches 293; Conservative 121; Mismatches 295; Indels 184; Gaps 18;
QY 65 YKTLFVEFTDHLFENIAKPRPPMGLGPTIOAEVYDTVITLKNMASHPVSLHAVGSY 124
Db 13 YKTIYKEYKDDSTDEVAQPAWGLFGLVLAEGVDVILHLKNPATRPTIHPHGVY 72
QY 125 WKASEGAEYDDQTSQREKEDDKVFPBGSHTYVQVLKENGPMASDPLCLTYSYLSHVDLV 184

Db 73 EKDSEGLYPDGSGLPKADSDVPPGSGHITNWTIPEGHAPTADAPACLWIIYHSHVDAP 132
QY 185 KDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFEDEKSWH-----SETKNSLMOD 238
Db 133 RDIATGLIGPLITCKRG-----TDHDFLLFSVDENLSWHLNENIATFYCSDPASVD 184
QY 239 RDAASARAWPKMHTVNGVYNSLPGILGCHRRKSVYWHVIGMTTPEVHSIFLEGHTFLVR 298
Db 185 KEDETFOESNRHAINGFVFGULPELNMCAQKRAVWHLFGMGNEIDVHTAFHGMQLTFR 244
QY 299 NHRQASLEISPTFLTAQTLMDLGQFLFCHIISSHQHDGMEAYVKVDSCEEPOLRMKN 358
Db 245 GHHTDVANIFPATFTVAEMVPEPBTWLISCQVNSHFRDGMQALYKVKSCSMAAPV---- 301
QY 359 NEEAEDYDDDLTDSMDVVRFDNNSPSFIQIRSVAKKHPRKTWVHYIAAEEDWDYAPL- 418
Db 301 -----DLTGT-----KVRQYFIEAHEIOWDYGPWG 325
QY 418 -----VLAPDRSYKSQYLNNGPQRIGRKYKVRFMAYTDEFTKTRALIOHES--GI 467
Db 326 HDGSTGKNLREP--SISDKFQKSSSRIGGYWKVRYEAQDETFOEKHLEEDRHLGI 383
QY 468 LGPRLXGEVGDITLILFKNOASRPYNIYPHGITDVRPLYSRLPKGVKHLKDPILPGEI 527
Db 384 LGPIRAEVGDTIQVVEYNRASQDFSMQPHGV-----FYEKDEGTV--YNGSSSPGLV 436
QY 528 FK-----YKWTYVEDGPTKSDPRCLTRYYSFVNMERDLASGLIGPLLYCKESVDOR 581
Db 437 AKPFEKVTYRWYVPHAGPTAODPACLTWMYFSAADPIRDTNSGLVGPLVCRAGALGAD 496
QY 582 GNQMSDKRNVLFSVEDENRSWYLTENIQRFLENPAGVQL-----EDPE-FOASNIMH 634
Db 497 GKQGVDKFEFLFTVLDENKSWYSN-----ANQAAAMLDFRLLSEDIQFQDSNRMH 549
QY 635 SINGYVFDLS-QLSVCLHEVAYWYLLSIGAQTDLSVFFSGYTFPKHMYEDTLLFPFS 693
Db 550 AINGFLFSNLPRLDMCKGDTVAWMLLGLGETVDVGMFQGNVQLQGRKGAAMLEPHT 609
QY 694 GETVFSMENPGWLILGCHNSDFRNKGMTALLKYSSCDKNTGDYEDSYEDISAYLLSKN 753
Db 610 FVMALQPDNLGTFELYCQAGSHREGAMRAIYVNSQC----- 647
QY 754 NAIEPRFSQNPVLRHOREITRTTLQSDQEIDYDDTISVEMKKEDEDIYDEDENQSP 813
Db 647 -----PGHQ-----ATP 653
QY 814 RSFOKTRHYFIAAVERLMDYGMSSS-PHYLRNRAQSGSV-----POFKKV 859
Db 654 RQRYQARIYYIMAEEVEMDYCPDRSWEREMHNSQEKDSYGYIFLSNKDGLGSRYKRAV 713
QY 860 FOEFTDGSFTQPLYRGELNEHLGLGPYIRAEVEDNINVTFRNQAASRPYSFYS 912
Db 714 FREYTDGTFRIPRPRGPEEHLGLIGPLIKGEVDILTVEFKNNASRPYSVHA 766

RESULT 8
ID 014286 PRELIMINARY; PRT; 216 AA.
AC 014286;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE COAGULATION FACTOR VIII ASSOCIATED PROTEIN B.
GN F8B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93052386.
RA LEVINSON B., KENWICK S., GAMEL P., FISHER K., GITSCHER J.;
RT "Evidence for a third transcript from the human factor VIII gene."
Genomics 14:585-589(1992).

077718		
ID	077718	PRELIMINARY; PRT; 363 AA.
AC	077718;	
DT	01-NOV-1998 (TREMBLrel. 08, Created)	
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)	
DT	01-MAY-1999 (TREMBLrel. 10, Last annotation update)	
DE	SPERM-MEMBRANE ASSOCIATED PROTEIN P47 (FRAGMENT).	
OS	Equus caballus (Horse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
OC	Eutheria; Perissodactyla; Equidae; Equus.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=TESTIS;	
RA	GENTZEL M., TOEPFER-PETERSEN E.;	
RL	Submitted (Aug-1998) to the EMBL/genBank/DBJ databases.	
DR	EMBL; AJ010121; CAA09010.1; -.	
DR	PFAM; PF00008; EGF; 1.	
DR	PFAM; PF00754; F5_F8_type_C; 2.	
FT	Sperm.	
FT	NON_TER 1	
FT	NON_TER 363	
SO	SEQUENCE 363 AA; 40744 MW; B470B46A CRC32;	

Query Match	6.6%;	Score 514.5;	DB 6;	Length 363;
Best Local Similarity	36.0%;	Pred. No. 3.5e-31;		
Matches 126;	Conservative 49;	Mismatches 118;	Indels 57;	Gaps 11;

```

0Y 1124 ECLLGEHLHAGMSTLELVY-----SNKCQTPLGMSGHIRDFQITASGOY-- 1169
    | : : | | : | | | | | | | | | | | | | | | |
Db 42 ECHVIDDSHRG--DVFTQYICSCPBGYTGTHCETTCAMPLGETGAIADAQISASSVYFG 99
    | : : | | : | | | | | | | | | | | | | | | |
0Y 1169 ---GOWAPKLARLHYSGSINAWST---KEPESWIKVDLLAPMIHGIKTGAROKFSS 1220
    : | : | | | : : : | | : | : | : | : | : | : : :
Uu 100 FMGLGRWVPELARLHRTGIVNAWMTASNYDKNP--WIQVNI MRKMRVTGVTTCASRGTA 157
    | : : | | : | : : | : | : | : | : | : | : | : :
0Y 1221 LYISQFLIMYSLDGKKWQTYRGNSHGTLMVEFGVNDSSGIKHNFNPIIARYIRLHPH 1280
    | : : | : | : | : | : | | | | | | : | : | : |
Db 158 EYLKTFKVAYSVDGRKFQIRDAGDSKDKVFVGANDNGLKVNHDVPLEVTVYRLVP-- 216
    | : : | : | : | : | : | | | | | | : | : | : |
0Y 1281 YSIRSTLRMELMGCDLNSCSMPLGESKAISDAQITASYE----TNMFATWSDSKARLH 1336
    : | : : | : | : | | | | | : | : | : | | |
Db 216 -----LACH-HGCTRP-----DRQITASSTYRTGLNAF-SWIFPYARLD 253
    | : : | : | : | : | : | : | : | : | : | : |
0Y 1337 LOGRSNAWRPQVNNPKEMLOVDFOQTKMVTGVTGQVKSLLTSMYVKEFLISSQDGHQW 1396
    | : | | | | : | : | | | : | | | | | : | : | : |
    254 KQGFENAWMTAOSNSASEMWLODDLGSQKEVTGVTITGARDFGHIQIYDAYKVSHSNDGANW 313
    | : | | | | : | : | | | : | : | : | : | : | : |
0Y 1397 TLFEPQ--GKVKVFQGDQDSFTPVVNSLDPPLTRRLRIHPQSWHQIAL 1444
    | : | : | | | : : | : | : | : | : | : | : |
Db 314 TEYRDQRAADSKIFLGNLDNNSHKKNMFTETPLAREVRILPVAMHRRITL 363

```

```

RESULT 12
060462
ID 060462 PRELIMINARY; PRT; 931 AA.
AC 060462:
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, last annotation update)
DE VASCULAR ENDOTHELIAL CELL GROWTH FACTOR 165 RECEPTOR 2.
GN VEGF165R2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BREAST:
RX MEDLINE; 98188099.
RA SOKER S., TAKASHIMA S., MIAO H.Q., NEUFELD G., KLAGSBRUN M.;
RT "Neuropilin-1 is expressed by endothelial and tumor cells as an
RT isoform-specific receptor for vascular endothelial growth factor.";
RL Cell 92:735-745(1998).

```

DR EMBL; AF016098; AAC12922.1; -.
DR PFAM; PF00431; CUB; 2.
DR PFAM; PF00754; F5_F8_type_C; 2.
DR PFAM; PF00629; MAM; 1.
SQ SEQUENCE 931 AA; 104830 MW; 912262F2 CRC32;

Query Match	6.08;	Score 469.5;	DB 4;	Length 931;
Best Local Similarity	28.58;	Pred. No. 4.4e-27;		
Matches 166; Conservative	87;	Mismatches 185;	Indels 145;	Gaps 32;

```
QY    953 DEFDOCK-AWAYFSDVDLEK---DVHSGLIGPLLVCHTNTL-----NPAHGROYTVQOFA 1002Z
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    79 EKHDCRYDFEIRPDGSEASADLGKHCNGNIAPTIISSGSMLYIKFTSDIARQA--GFS 136

QY    1003 LEFTTFDETksWYFTENMERNCRAPCNIQMEDPFEKENYRFRHAINGYIMDTLLGLVMAQD 1062Z
      |:|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    137 LRYELF-KTGS----EDCSKNFTSP-NGTIESPGEPEKP----- 171

QY    1063 QRIRWYLLSMGSNNENIHSHFSGHVFTVRKKKEEXKMALYNLPVGVEFTEVEMLSKAG--- 1120Z
      |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    171 -----HNLDCTFTILAKPKMEIILIQFL--IFD-LHHDDLQVGEGD 207
```

QY	1120	-----IW-----RVECLIGEH-----LHAGMST-----LFLV	1141
Db	208	CKYMDLIDWIGIPHNHGPPLIGKYCGKTPSELRSSTGILSLTFHDMAMAVAKDGFSAHYLYV	267
QY	1142	YSN-----KCQTPLCMASGHIRDPQITASGOY--GOWAPKLARLHYSGSINAW-----ST	1189
Db	268	HQEPLENFQCNPVPLGMEGRIANEJISASSTYSIGRWTPQOSRLH--GDDNGWTPNLDSDN	325
QY	1190	KEPFWSIKVDLAPMLIHGIKTQGA--ROKFSLSYISOFIIMYSLUDGKKWQTYRGNSTGT	1247
Db	326	KE---YLQVDLRFLTMLTATQGAISRETQNGYVKSYLEVSTNGEDMMVYRHGKNH--	382
QY	1248	LMVEFGNVDSSGIKHNIFNPPIIARYIRLHPTHYSIRSTLRMELMGCDLNS--CSMPLGM	1305
Db	382	-KVFQANNDATEEVLNKLHAPLITRFRVIRIPQWHSICIALRLELFGCRVTDAPCSNMLGM	440
QY	1306	ESKAISDAQITASSYFTNMFATWSPSKARLHLOGRSNW--RPOVNNPKEWLOVDFOKT	1362
Db	441	LSGLIADSQISASS--TOEY-LWSPSAARL-VSSRS-GWFPRIPOAQOGEELWLOVDLGT	495
QY	1363	MKVTVGTTQO-----VKSLLTSMYVKEFLISSODGHQWTLFONGKV--KYVQGNOD	1413
Db	496	KTVKGVILOGARGGDSITAVEARAFYRKFKVYSYSLNGKDWE-YODPRTQOPKLEEGNMH	554
QY	1414	SFTPVVNSLDPLLTRYLRIHPQSVWHQ--IALRMEVLGCEAOD	1455
Db	555	YDTPDIRRFD-PIPAQYVRYVPERWSPAGIGMRLEVLGCDWTD	596

RESULT	13		
ID	014820		
ID	014820	PRELIMINARY;	PRT; 909 AA.
AC	014820;		
DT	01-JAN-1998	(TREMBLrel. 05, Created)	
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)	
DT	01-MAY-1999	(TREMBLrel. 10, Last annotation update)	
DE	NEUROPIILIN-2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
CC	Eutheria; Primates; Catarrhini; Homiinae; Homo.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE; 97470888.		
RA	CHEN H., CHEBOTAL A., HE Z.-G., GOODMAN C.S., TESSIER-LAVIGNE M.;		
RT	"Neupopiln-2, a novel member of the neupopiln family, is a high		
RT	affinity receptor for the semaphorins Sema E and Sema IV but not Sema		
RL	III.";		
RL	Neuron 19:547-559(1997).		
DR	EMBL; AF022859; AAC51788.1; -.		
DR	PFAM; PF00431; CUB; 2.		

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 1999, 15:18:23 ; Search time 17.89 Seconds
(without alignments)
1929.047 Million cell updates/sec

Title: US-09-001-039a-47
Perfect score: 7797
Sequence: 1 MQIELSTCFFLCFLRRCFSA.....WVHQLALRMEVLGCEAODLY 1457

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	7797	100.0	1457	1	W44372	Human Factor VIII
2	7797	100.0	1457	1	W46246	Human Factor VIII
3	7759	99.5	1471	1	W23414	Human B-domain del
4	7668.5	98.4	1440	1	R12971	Factor VIII:SQ. Re
5	7620.5	97.7	1516	1	P80265	Modified factor VI
6	7590	97.3	1425	1	P80267	Modified factor VI
7	7587	97.3	1424	1	P80268	Sequence of 740 Ar
8	7583	97.3	1424	1	P91169	Factor VIII-dB695-
9	7544	96.8	1661	1	W18670	Procoagulant-activ
10	7362.5	94.4	1383	1	W33227	Procoagulant-activ
11	7354.5	94.3	1383	1	W33229	Procoagulant-activ
12	7351.5	94.3	1383	1	W33228	Procoagulant-activ
13	7344.5	94.2	2342	1	W11422	Active Factor VIII
14	7343.5	94.2	2344	1	W11432	Active Factor VIII
15	7343.5	94.2	2345	1	W11410	Active Factor VIII
16	7343	94.2	2345	1	W11415	Active Factor VIII
17	7343	94.2	2345	1	W11403	Active Factor VIII
18	7342.5	94.2	2346	1	W11421	Active Factor VIII
19	7342.5	94.2	2346	1	W11431	Active Factor VIII
20	7342.5	94.2	2346	1	W11434	Active Factor VIII
21	7342	94.2	2347	1	W11411	Active Factor VIII
22	7342	94.2	2347	1	W11402	Active Factor VIII
23	7341.5	94.2	2348	1	W11409	Active Factor VIII
24	7341.5	94.2	2348	1	W11414	Active Factor VIII
25	7341	94.2	2349	1	W11424	Active Factor VIII
26	7341	94.2	2349	1	W11428	Active Factor VIII
27	7341	94.2	2349	1	W11430	Active Factor VIII
28	7341	94.2	2349	1	W11420	Active Factor VIII
29	7341	94.2	2349	1	W11401	Active Factor VIII
30	7340.5	94.1	2350	1	W11405	Active Factor VIII
31	7340.5	94.1	2350	1	W11413	Active Factor VIII
32	7340	94.1	2351	1	P50059	Human factor VIII.
33	7340	94.1	2351	1	W00465	Factor-VIII. Retro
34	7340	94.1	2351	1	W11425	Active Factor VIII
35	7340	94.1	2351	1	W11427	Active Factor VIII
36	7340	94.1	2351	1	W11435	Active Factor VIII
37	7340	94.1	2351	1	W11408	Active Factor VIII
38	7340	94.1	2351	1	W11416	Active Factor VIII
39	7340	94.1	2351	1	W11419	Active Factor VIII
40	7340	94.1	2351	1	W11398	Active Factor VIII
41	7340	94.1	2351	1	W11399	Active Factor VIII
42	7340	94.1	2351	1	W11404	Active Factor VIII
43	7340	94.1	2351	1	W11329	Native Factor VIII

ALIGNMENTS

RESULT	1	
W44372	standard; Protein: 1457 AA.	
ID	W44372	
AC	W44372;	
DT	20-JUL-1998 (first entry)	
DE	Human Factor VIII SQN deletion mutant.	
KW	Factor VIII; blood clotting; haemophilia A; gene therapy;	
KW	retrovirus; vector; human.	
OS	Homo sapiens.	
OS	Synthetic.	
PN	WO9800542-A2.	
PD	08-JAN-1998.	
PE	02-JUL-1997; U11785.	
PR	04-JUN-1997; US-869309.	
PR	03-JUL-1996; US-645601.	
PR	13-AUG-1996; US-696381.	
PA	(CHIR) CHIRON CORP.	
PI	Allen JR, Barber JR, Boder M, Chang SM, Chong K,	
PI	De La Vega D, Depolo NJ, Greengard J, Hsu DC, Ibanez CE,	
PI	Jolly DJ, Mittelstaedt DM, Prussak CE, Respass JG,	
DR	WPI: 98-086967/08.	
DR	N-PSDB: V15338.	
PT	New replication defective recombinant retroviruses - which express B	
PT	domain-deleted human factor VIII or human factor IX for the	
PT	treatment of haemophilia	
PS	Claim 5: Page 175-180; 236pp; English.	
CC	This polypeptide comprises the B domain deletion mutant SQN of	
CC	human Factor VIII. The SQN mutant is created by fusing Ser-743 to	
CC	Gln-1638 of native Factor VIII (see W44373) to form a Ser-Gln-Asn	
CC	(SQN) link between the A2 and A3 Factor VIII domains. A DNA	
CC	sequence encoding the SQN deletion mutant is provided in V15338.	
CC	When compared to plasmid-derived Factor VIII, the SQN deletion does	
CC	not influence the in vivo pharmacokinetics, but the reduced size of	
CC	the molecule appears to decrease proteolytic degradation. The	
CC	invention relates to preparations of replication defective	
CC	recombinant retrovirus (RV) expressing a B domain-deleted human	
CC	Factor VIII protein, where the recombinant RV is capable of	
CC	infecting human cells, is resistant to degradation by human	
CC	complement and is capable of inducing long-term (at least 30 days	
CC	and up to 6 months or longer post-injection) systemic expression of	
CC	Factor VIII when administered to a haemophilia A patient.	
SQ	Sequence 1457 AA;	
Query Match 100.0%; Score 7797; DB 1; Length 1457;		
Best Local Similarity 100.0%; Pred. No. 0;		
Matches 1457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 MQIELSTCFCLLRFCSATRRYYLGAVELSDYMQSDIGELPVDARFPPRPKSPFN 60	
DB	1 MQIELSTCFCLLRFCSATRRYYLGAVELSDYMQSDIGELPVDARFPPRPKSPFN 60	
OY	61 TSVVYKKTLEFEFTDHLFNIAKPRPMMGLGPTIQAEVYDVTVTLLKNMASHPVLHAV 120	
DB	61 TSVVYKKTLEFEFTDHLFNIAKPRPMMGLGPTIQAEVYDVTVTLLKNMASHPVLHAV 120	
OY	121 GVSYWKASEGAEYDDQTSQREKEDDKVFPGSHTYVWQVLKENGPMASDPLCLTYSLSH 180	
DB	121 GVSYWKASEGAEYDDQTSQREKEDDKVFPGSHTYVWQVLKENGPMASDPLCLTYSLSH 180	
OY	181 VDLVKDLSGLIGALLVCRESSLAKEKQTLLHKFILLFAVDEGKSWHSEKNSLMQDRD 240	
DB	181 VDLVKDLSGLIGALLVCRESSLAKEKQTLLHKFILLFAVDEGKSWHSEKNSLMQDRD 240	
OY	241 AASARAWPKMHTVNGYVNRSLPGLIGCHRSYVWHVIGMGTTPPEVHSIFLEGHTFLVRNH 300	

Db 241 AASARAWPKMHTVNGVYVNRSLPGLIGCHRRKSVYWHVIGMGTPEVHSIFLESGHTELVNRH 300
 QY 301 ROASLEISPTITLTAOTLMDLGOFLFCHISSHQHDMGEAYKVDSCEPEOLRMKNNE 360
 Db 301 ROASLEISPTITLTAOTLMDLGOFLFCHISSHQHDMGEAYKVDSCEPEOLRMKNNE 360
 QY 361 EAEDYDDDLTDEMDVVRFDNDSPSFIQIRSVAKKHPTWVYIAAEEDWDYAPLVLA 420
 Db 361 EAEDYDDDLTDEMDVVRFDNDSPSFIQIRSVAKKHPTWVYIAAEEDWDYAPLVLA 420
 QY 421 PDDRSYKSOYLNNGPQRIGRKYKVFEMAYTDETFKTRERAIQHSIGLPLLYGEVGDJL 480
 Db 421 PDDRSYKSOYLNNGPQRIGRKYKVFEMAYTDETFKTRERAIQHSIGLPLLYGEVGDJL 480
 QY 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIEFKYKWTVEDGP 540
 Db 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIEFKYKWTVEDGP 540
 QY 541 TKSDRCLTRYSSSFVNMERDLASGLIGPLLYCYKESVDQRGNOIMSDKRNVLFSVDE 600
 Db 541 TKSDRCLTRYSSSFVNMERDLASGLIGPLLYCYKESVDQRGNOIMSDKRNVLFSVDE 600
 QY 601 NRSWYLTENIQRFLPNPAGVOLLEDPFQASNIMHSINGYVDSIOLSVCLHEVAYWYLS 660
 Db 601 NRSWYLTENIQRFLPNPAGVOLLEDPFQASNIMHSINGYVDSIOLSVCLHEVAYWYLS 660
 QY 661 IGAQDFLSVFFSGYTFKHMYEDTLTLPFSGEYVEMSMENPGMILGCHNSDFRNRG 720
 Db 661 IGAQDFLSVFFSGYTFKHMYEDTLTLPFSGEYVEMSMENPGMILGCHNSDFRNRG 720
 QY 721 MTALLKYSSCDKNTGDYEDSYEDISAYLLSKNNAIEPRFSQNPVLPKRHQREITRTTL 780
 Db 721 MTALLKYSSCDKNTGDYEDSYEDISAYLLSKNNAIEPRFSQNPVLPKRHQREITRTTL 780
 QY 781 QSDQEEIDYDDTISVEMKKEDEFDIYDEDENSPSRFOKTRHYFLAVERLMDYGMSSSP 840
 Db 781 QSDQEEIDYDDTISVEMKKEDEFDIYDEDENSPSRFOKTRHYFLAVERLMDYGMSSSP 840
 QY 841 HVLNRRAQSGSVPOFKKVVQOEFTDGSFTQPLYRGELNEHLGLGPYIRAEVEDNINWTF 900
 Db 841 HVLNRRAQSGSVPOFKKVVQOEFTDGSFTQPLYRGELNEHLGLGPYIRAEVEDNINWTF 900
 QY 901 RNQASRPYSFYSSLISYEEDQROGAEPKRNFKVKNETTYFWKVQHMAPTKDEFDCKAW 960
 Db 901 RNQASRPYSFYSSLISYEEDQROGAEPKRNFKVKNETTYFWKVQHMAPTKDEFDCKAW 960
 QY 961 AYSFSDVLEKDVHSGLIGPLLVCHTNTLPAHGRQVTVQEFALFTTIDETKSWYFTENA 1020
 Db 961 AYSFSDVLEKDVHSGLIGPLLVCHTNTLPAHGRQVTVQEFALFTTIDETKSWYFTENA 1020
 QY 1021 ERNCRAPCNIQMEDPTFEKENVRFHAINGYTMDTLPGLVNAQDQIRRWYILSMGSNENIHS 1080
 Db 1021 ERNCRAPCNIQMEDPTFEKENVRFHAINGYTMDTLPGLVNAQDQIRRWYILSMGSNENIHS 1080
 QY 1081 IHFSGHVFTVRKKKEEYKMAALYNLYPGVEFEYEMLPKAGIWRVECLIGEHLHAGMSTLFL 1140
 Db 1081 IHFSGHVFTVRKKKEEYKMAALYNLYPGVEFEYEMLPKAGIWRVECLIGEHLHAGMSTLFL 1140
 QY 1141 VYSNKCQPTLGMASGHIRDPQITASGQYGGWAPKLARLHYSGSINAWSTKEPFSWIKVDL 1200
 Db 1141 VYSNKCQPTLGMASGHIRDPQITASGQYGGWAPKLARLHYSGSINAWSTKEPFSWIKVDL 1200
 QY 1201 LAPMIHIGITOGAROKFSSLYISOFIIMYSLDGKKMOTYRGNSTGTLMVFFGNVDSGI 1260
 Db 1201 LAPMIHIGITOGAROKFSSLYISOFIIMYSLDGKKMOTYRGNSTGTLMVFFGNVDSGI 1260
 QY 1261 KHNIFNPITARIYIRLHPTHYISIRSTLRMELMGCDLNSCMBPLGMEKASIDAQITASSY 1320
 Db 1261 KHNIFNPITARIYIRLHPTHYISIRSTLRMELMGCDLNSCMBPLGMEKASIDAQITASSY 1320
 QY 1321 FTNMFATWSPSKARLHLQGRSNAMRPQVNNPKEMLOVDFOKIMKVTGVTTOGVKSLLTSM 1380
 Db 1321 FTNMFATWSPSKARLHLQGRSNAMRPQVNNPKEMLOVDFOKIMKVTGVTTOGVKSLLTSM 1380

QY 1381 YKKEFLISSQDGHQWTLFFQNGKYKVFQGNODSFPTPVNSLDPPLLTRYLRIHQSWVH 1440
 Db 1381 YKKEFLISSQDGHQWTLFFQNGKYKVFQGNODSFPTPVNSLDPPLLTRYLRIHQSWVH 1440
 QY 1441 QIALRMEVYLGCSEAQDLY 1457
 Db 1441 QIALRMEVYLGCSEAQDLY 1457
 RESULT 2
 ID W46246 standard; Protein: 1457 AA.
 AC W46246;
 DE 06-AUG-1998 (first entry)
 KW Human factor VIII beta-domain deleted SQN deletion protein sequence.
 KW Replication defective; recombinant retrovirus; RRV; therapeutic protein;
 KW haemophilia; thrombosis; hypercoagulable disorder; liver disease; human;
 KW hepatitis; thalassemia; phenylketonuria; Lesch-Nyhan syndrome; diabetes;
 KW cystic fibrosis; Duchenne's Muscular Dystrophy; hypercholesterolemia;
 KW hypopituitarism; adenine deaminase deficiency; HIV infection; anaemia;
 KW Guacher's syndrome; high blood pressure; Alzheimer's disease; autoimmune;
 KW inflammatory disease; factor VIII.
 OS Homo sapiens.
 PN WO9800541-A2.
 PD 08-JAN-1998.
 PF 02-JUL-1997: U11784.
 PR 04-JUN-1997: US-869309.
 PR 03-JUL-1996: US-645601.
 PR 13-AUG-1996: US-696381.
 PA (CHIR) CHIRON CORP.
 PI Allen JR, Barber JR, Boder M, Chang SMW, Chong K,
 PI De LA VEGA D, Depolon J, Greengard J, Hsu DC, Ibanez CE,
 PI Jolly DJ, Lee R, Mittelstaedt DM, Prussak CE, Respass JG;
 DR WPI: 98-086966/08.
 DR N-PSDB: V19581.
 PT New replication defective recombinant retro-viruses - which can be
 PT administered to provide long term systemic expression of therapeutic
 PT protein in blood, useful in, e.g. treating hyper-coagulable
 PT disorders
 PS Example 28; Pages 213-217; 272pp; English.
 CC This is the beta-domain deleted SQN deletion protein of human factor
 CC VIII. The encoding DNA is used to construct recombinant retroviral
 CC vectors expressing human factor VIII. The invention provides the
 CC preparation of replication defective recombinant retrovirus (RRV)
 CC expressing a therapeutic protein. The RRV preparation is resistant to
 CC degradation by human complement and is capable of inducing long term
 CC systemic expression of the therapeutic protein when administered
 CC intravenously to a human. The long term systemic expression results in a
 CC measurable level of the therapeutic protein being produced in the blood
 CC of the human for a period of at least 30 days after the administration of
 CC the RRV vector preparation. RRV's can be used for in vivo delivery of
 CC therapeutic protein to treat, e.g. haemophilia A, haemophilia B,
 CC thrombosis, hypercoagulable disorders, liver diseases such as hepatitis,
 CC disorders such as thalassemia, phenylketonuria, Lesch-Nyhan syndrome,
 CC severe combined immunodeficiency (SCID), cystic fibrosis, Duchenne's
 CC Muscular Dystrophy, inherited emphysema, familial hypercholesterolemia,
 CC diabetes, hypopituitarism, adenine deaminase deficiency, alpha1-
 CC antitrypsin deficiency, Guacher's syndrome, anaemia, infections such as
 CC HIV infection, high blood pressure, Alzheimer's disease, autoimmune or
 CC inflammatory disease or graft versus host disease. RRV's are capable of
 CC surviving inactivation in human serum thereby allowing efficient gene
 CC transfer over prolonged periods of time.
 SQ Sequence 1457 AA;
 Query Match 100.0%; Score 7797; DB 1; Length 1457;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MOIELSTCFELCLIRCFESATRRYYLGAVELSWDMQSDLGELPYDARFPPRVKSPFN 60
 Db 1 MOIELSTCFELCLIRCFESATRRYYLGAVELSWDMQSDLGELPYDARFPPRVKSPFN 60

```
QY 61 TSVVYKKTLEVEFTDHLFNIAKPRPMPMGLGPTIOAEVYDTVITLKNMASHPVSLHAV 120
DB 61 TSVVYKKTLEVEFTDHLFNIAKPRPMPMGLGPTIOAEVYDTVITLKNMASHPVSLHAV 120
QY 121 GVSVMKASEGAEDDQTSOREKEDDKVPPGSGSHTYVMQVLKENGPMASDPLCLTYSLSH 180
DB 121 GVSVMKASEGAEDDQTSOREKEDDKVPPGSGSHTYVMQVLKENGPMASDPLCLTYSLSH 180
QY 181 VDLVKDNLNGLIGALLVCREGSLAKEKTQTLHKFILLEAVFDEGKSWHSETKNSLMQDRD 240
DB 181 VDLVKDNLNGLIGALLVCREGSLAKEKTQTLHKFILLEAVFDEGKSWHSETKNSLMQDRD 240
QY 241 AASARAWPKMHTVNGVYVNRSLPGLIGCHRSKYVMHVIGMGTTPVHSTFLEGTFLVNRH 300
DB 241 AASARAWPKMHTVNGVYVNRSLPGLIGCHRSKYVMHVIGMGTTPVHSTFLEGTFLVNRH 300
QY 301 RQASLEISPTFTLTAQTLMLMDLGQFLFCHISSHQHDMGMEAYVKVDSCEPEPQLRMKNNE 360
DB 301 RQASLEISPTFTLTAQTLMLMDLGQFLFCHISSHQHDMGMEAYVKVDSCEPEPQLRMKNNE 360
QY 361 EAEDYDDDLTDEMDVYRFDDDNPSFTQIRSVAKKHPTWVHYIAEEDWDYAPLVLA 420
DB 361 EAEDYDDDLTDEMDVYRFDDDNPSFTQIRSVAKKHPTWVHYIAEEDWDYAPLVLA 420
QY 421 PDDRSYKSOYLNNGPQRIGRKYKRVKRMAYTDEFKTRFAIOHESGILGPLYGEVGTDL 480
DB 421 PDDRSYKSOYLNNGPQRIGRKYKRVKRMAYTDEFKTRFAIOHESGILGPLYGEVGTDL 480
QY 481 LIIFKNOASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIRKYKWTVTVEDGP 540
DB 481 LIIFKNOASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIRKYKWTVTVEDGP 540
QY 541 TKSDPRCLTRYSSSFVNMERDLASGLIGPLLICYESVDQRGNOIMSDKRNVLFSYVDE 600
DB 541 TKSDPRCLTRYSSSFVNMERDLASGLIGPLLICYESVDQRGNOIMSDKRNVLFSYVDE 600
QY 601 NRSWYLTENIORFLPNAGVQLEDPEDQASNMHMSINGYVDSLOLSVCLHEVAYWYLLS 660
DB 601 NRSWYLTENIORFLPNAGVQLEDPEDQASNMHMSINGYVDSLOLSVCLHEVAYWYLLS 660
QY 661 IGAQTFDLVSFFSGYTFKHKMYEDTLTFPFSGETVMSMENPGMLWILGCHNSDFRNRG 720
DB 661 IGAQTFDLVSFFSGYTFKHKMYEDTLTFPFSGETVMSMENPGMLWILGCHNSDFRNRG 720
QY 721 MTALKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRFSQNPVLRKHOREITRTTL 780
DB 721 MTALKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRFSQNPVLRKHOREITRTTL 780
QY 781 QSDQEEIDYDDTISVENMKEDFDIYEDENQSPRSFQKTRHYFIAAVERLMDYGMSSSP 840
DB 781 QSDQEEIDYDDTISVENMKEDFDIYEDENQSPRSFQKTRHYFIAAVERLMDYGMSSSP 840
QY 841 HVLNRRAQSGSVPOFKVYFOEFTDGSFTQPLRGELNEHLGLGPYIRAEVEDNIMYTF 900
DB 841 HVLNRRAQSGSVPOFKVYFOEFTDGSFTQPLRGELNEHLGLGPYIRAEVEDNIMYTF 900
QY 901 RNOASRPYSFYSSLISYEEDQROGAEPKRNFKVKNETKTYFWKVQHMAHPKDEFDCKAW 960
DB 901 RNOASRPYSFYSSLISYEEDQROGAEPKRNFKVKNETKTYFWKVQHMAHPKDEFDCKAW 960
QY 961 AYSVDVLEKDVHSGLIGPLLVCHTNTLNPAGROVTVQEFALFTTFDETKSWYFTENM 1020
DB 961 AYSVDVLEKDVHSGLIGPLLVCHTNTLNPAGROVTVQEFALFTTFDETKSWYFTENM 1020
QY 1021 ERNCRAPCNQIOMEDPTFKENYRFHAINGYIMDTLPGLVMAQDORIRWYLLSMGSNENIHS 1080
DB 1021 ERNCRAPCNQIOMEDPTFKENYRFHAINGYIMDTLPGLVMAQDORIRWYLLSMGSNENIHS 1080
QY 1081 IHFSGHVFVRKKEEYKMALYNLYPGVFTEVEMLPKAGIMRVECLIGELHAGMSTLFL 1140
DB 1081 IHFSGHVFVRKKEEYKMALYNLYPGVFTEVEMLPKAGIMRVECLIGELHAGMSTLFL 1140
```

```
QY 1141 VYSNRCQTPLGMAHGHIRDFOITASGOYGOWAPKLARLHSGSINAMSTKEPFSWIKVDL 1200
DB 1141 VYSNRCQTPLGMAHGHIRDFOITASGOYGOWAPKLARLHSGSINAMSTKEPFSWIKVDL 1200
QY 1201 LAPMIHGIKTQAGAROFSSLYISOFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSGSI 1260
DB 1201 LAPMIHGIKTQAGAROFSSLYISOFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSGSI 1260
QY 1261 KHNIFNPPIIARIYRHPHTYSIRSRTLMEMLGCDLNSCMLPGMESKAISDAQITASSY 1320
DB 1261 KHNIFNPPIIARIYRHPHTYSIRSRTLMEMLGCDLNSCMLPGMESKAISDAQITASSY 1320
QY 1321 FTNMFATWSPSKARLHLQGRSNAMRPQVNNPKEWLQVDFOKTMYTGVTTOGVKSLTSM 1380
DB 1321 FTNMFATWSPSKARLHLQGRSNAMRPQVNNPKEWLQVDFOKTMYTGVTTOGVKSLTSM 1380
QY 1381 YVKEFLISSQDGHQWTLFFONGKYVFOGNQDSFTPVVNSLDPPLLTRYLRHHPQSWVH 1440
DB 1381 YVKEFLISSQDGHQWTLFFONGKYVFOGNQDSFTPVVNSLDPPLLTRYLRHHPQSWVH 1440
QY 1441 QIALRMEVLGCEAODLY 1457
DB 1441 QIALRMEVLGCEAODLY 1457

RESULT 3
ID W23414 standard; Protein; 1471 AA.
AC W23414;
DT 08-APR-1998 (first entry)
DE Human B-domain deleted factor VIII protein.
KW Post-translational regulatory element; PRE; enhancer II; intronless gene;
KW surface antigen gene; cytoplasmic accumulation; targeted delivery;
KW near consensus splice sequence; blood coagulation factor; factor VIII;
KW factor IX.
OS Homo sapiens.
PN MO9733994-A1.
PD 18-SEP-1997.
PF 10-MAR-1997; U03561.
PR 11-MAR-1996; US-683839.
PA (IMMU-) IMMUNE RESPONSE CORP.
PI Bidlingmayer S, III CR;
DR MPI; 97-470874/43.
DR N-PSDB; T73164.
PT Vector for increased expression of intronless genes - comprises
PT intronless gene with at least one near consensus splice sequence, a
PT promoter and at least one viral cis-acting post-transcriptional
PT regulatory element
PS Example 1; Pages 31-36; 59pp; English.
CC The present sequence represents human B-domain deleted factor VIII
CC The cDNA encoding this protein also contains, 3' of the coding region, a
CC post-translational regulatory element (PRE) of the Hepatitis B virus.
CC PRE sequences have been shown to function in cis to increase the
CC steady-state levels of surface gene transcripts by facilitating the
CC cytoplasmic accumulation of these transcripts. The above nucleic acid
CC sequence is part of a novel vector, comprising an intronless gene
CC containing 1 or more near consensus splice sequences operably linked to
CC a promoter sequence so that the gene is transcribed in a cell.
CC Intronless gene transcripts which contain near consensus splice site
CC sequences are believed to get tied up in the nucleus of the cell where
CC splicing occurs, rather than being transported to the cytoplasm where
CC they can be translated into proteins. The PRE sequences are transcribed
CC along with the gene, causing export of the gene transcript from the
CC nucleus into the cytoplasm of the cell. The vector can be used
CC to increase the expression of an intronless gene containing at least one
CC near consensus splice sites, preferably cDNA encoding a blood coagulation
CC factor, particularly factor VIII or IX. The complex allows the targeted
CC delivery of the vector to a specific cell, e.g. hepatocytes when the
CC ligand is an asialoglycoprotein which binds the asialoglycoprotein
CC receptor present on their surface.
SQ Sequence 1471 AA;
```

Query Match	99.5%;	Score 7759;	DB 1;	Length 1471;
Best Local Similarity	98.8%;	Pred. No. 0;		
Matches 1454; Conservative	0;	Mismatches 3;	Indels 14;	Gaps 1;

QY	1	MQIEJSTCFEFLCLLRFCFSATRRYYLGAVELSWDYMOSDLGELPVDAREPPRPVKSEFPEN	60
Db	1	MVEJELSTCFELCLLRFCFSATRRYYLGAVELSWDYMOSDLGELPVDAREPPRPVKSEFPEN	60
QY	61	TSVVYKKTLEVEFTDHLFNIAKPRPMPMGLGPITQAEVYDVTVITLKMAASHPVSLHAIV	120
Db	61	TSVVYKKTLEVEFTVHLFNIAKPRPMPMGLGPITQAEVYDVTVITLKMAASHPVSLHAIV	120
QY	121	GVSYWKASGAEYDDQTSOREKEDDKVFPGGSHTYYWQVLKENGPMASDPLCLTYSLSH	180
Db	121	GVSYWKASGAEYDDQTSOREKEDDKVFPGGSHTYYWQVLKENGPMASDPLCLTYSLSH	180
QY	181	VDLVKDLNSGLIGALLVCNREGSLAKEKTQTLHKFILLFAVEDEGKSWHSETKNSLMODRD	240
Db	181	VDLVKDLNSGLIGALLVCNREGSLAKEKTQTLHKFILLFAVEDEGKSWHSETKNSLMODRD	240
QY	241	AASARAWPKMHTVNGYVNRSLPGLIGCHRSYVWHVIGMGTPEVHSIFLEGHTEFLVRNH	300
Db	241	AASARAWPKMHTVNGYVNRSLPGLIGCHRSYVWHVIGMGTPEVHSIFLEGHTEFLVRNH	300
QY	301	ROASLEISBITFELTAQTLMDLGOFLLFCHTSSHODGMEAYVKVDSCEPEBOLRMKNNE	360
Db	301	ROASLEISBITFELTAQTLMDLGOFLLFCHTSSHODGMEAYVKVDSCEPEBOLRMKNNE	360
QY	361	EAEDYDDDLTDESEMDVVRFPDDNSPSFIQIRSAVAKKHPTWVHIAAEEDWDIAPLVLA	420
Db	361	EAEDYDDDLTDESEMDVVRFPDDNSPSFIQIRSAVAKKHPTWVHIAAEEDWDIAPLVLA	420
QY	421	PDDRSYKSOYLNGPQIRGRYKKVRFMAYTDETEKTRREALQIHESGILGPLLYGEVDTL	480
Db	421	PDDRSYKSOYLNGPQIRGRYKKVRFMAYTDETEKTRREALQIHESGILGPLLYGEVDTL	480
QY	481	LIIFKNOASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIIFYKKWTVTEDGP	540
Db	481	LIIFKNOASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIIFYKKWTVTEDGP	540
QY	541	TKSDPRCLTRYSSFVNMERDIASGLIGPLLICYKESVDQRGNOIMSDKRNVLFSVFDE	600
Db	541	TKSDPRCLTRYSSFVNMERDIASGLIGPLLICYKESVDQRGNOIMSDKRNVLFSVFDE	600
QY	601	NRSWYLTENIQRFLPNPAGVOLDEPFEQASNMHSINGYVFDLSLOLSVCLHEVAIWTILS	660
Db	601	NRSWYLTENIQRFLPNPAGVOLDEPFEQASNMHSINGYVFDLSLOLSVCLHEVAIWTILS	660
QY	661	IGAQTDFLSVFSFGYTFKHKMYEDTLTLFPFSGETVFMSEMPGLMILGCHNSDFRNRG	720
Db	661	IGAQTDFLSVFSFGYTFKHKMYEDTLTLFPFSGETVFMSEMPGLMILGCHNSDFRNRG	720
QY	721	MTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSFSON-----PP	766
Db	721	MTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSFSON-----PP	766
QY	767	VLKRHOREITRTTLQSDQOEIYDDTISVEMKKEDDIYEDENOSPRSFOKKTTRHYFIA	826
Db	767	VLKRHOREITRTTLQSDQOEIYDDTISVEMKKEDDIYEDENOSPRSFOKKTTRHYFIA	826
QY	827	AVERLMDYGMSSSPHVLNRNAQSGVPOFKKVVOEFTDGSFTQPLRYGELNEHLGLIGP	886
Db	827	AVERLMDYGMSSSPHVLNRNAQSGVPOFKKVVOEFTDGSFTQPLRYGELNEHLGLIGP	886
QY	887	YIRAEVEDNIMVTFERNQASRPYSFYSSLSIYEEDQOGAEPKKNFVKPNETKTYFWKYQH	946
Db	887	YIRAEVEDNIMVTFERNQASRPYSFYSSLSIYEEDQOGAEPKKNFVKPNETKTYFWKYQH	946
QY	947	HMAPTKDEFDCAKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAGROVTVQEFALFTT	1006
Db	947	HMAPTKDEFDCAKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAGROVTVQEFALFTT	1006
QY	1007	IPDEFKSWYFTTENMERNCRAPCNIQMEDPTFEKENYREHAINGYIMDTLGLVMAQDQIR	1066

Db	1021	IFDETKSWYFIENMERNCRAPCNIQMEDPTFKENYRFHANGYIMDTLPGLVMAQDQIR	1080
Qy	1067	WYLLSMGSNENHSHIHFSGHFTVRKKKEEKKMALLNLYPGVEFTEVMBPSKAGIWRVECL	1126
Db	1081	WYLLSMGSNENHSHIHFSGHFTVRKKKEEKKMALLNLYPGVEFTEVMBPSKAGIWRVECL	1140
Qy	1127	IGEHLHAGMSTLEFLVYSNKCOTPLGMA SGHIRDFQITASQYGOWAPKRLRHYSGSINA	1186
Db	1141	IGEHLHAGMSTLEFLVYSNKCOTPLGMA SGHIRDFQITASQYGOWAPKRLRHYSGSINA	1200
Qy	1187	WSTKEPEFSWIKYDLLAPMIHGIKTQGAROKFSSLYISQPIIMYSLDGKKWQTYRGNSIG	1246
Db	1201	WSTKEPEFSWIKYDLLAPMIHGIKTQGAROKFSSLYISQPIIMYSLDGKKWQTYRGNSIG	1260
Qy	1247	TLWVEFGNVDSGCIKHNIENPPIIARYIRLHPHYSIRSTLRMELMGCDLNSCSMPLGME	1306
Db	1261	TLWVEFGNVDSGCIKHNIENPPIIARYIRLHPHYSIRSTLRMELMGCDLNSCSMPLGME	1320
Qy	1307	SKAISDAOITASSYFTNMFATWSPSKARLHOGRSNAWRPOVNNPKEWLOVDFOKTMKVT	1366
Db	1321	SKAISDAOITASSYFTNMFATWSPSKARLHOGRSNAWRPOVNNPKEWLOVDFOKTMKVT	1380
Qy	1367	GVTTQGVKSLTSMYVKEFLISSQDGHOWTLFFQNGKVKYFQGNQDSFTPVVNSLDPL	1426
Db	1381	GVTTQGVKSLTSMYVKEFLISSQDGHOWTLFFQNGKVKYFQGNQDSFTPVVNSLDPL	1440
Qy	1427	LTRYLRIPHOSWVHQAIRMEVLGCEAODLY 1457	
Db	1441	LTRYLRIPHOSWVHQAIRMEVLGCEAODLY 1471	

RESULT 4

ID R12971 standard; protein; 1440 AA.

AC R12971;

DT 02-OCT-1991 (first entry)

DE Factor VIII:SQ.

KW Factor VIII; B domain; haemophilia.

PN WO9109122-A.

PD 27-JUN-1991.

PF 06-DEC-1990; SE0809.

PR 15-DEC-1989; SE-004239.

PA (KAB) KABIIVTRUM AB.

PI Almedstedt AB, Hellestrom EM, Larsson K, Lind P, Sandberg HI, Spira J, Sydow-Backman M;

PI WPI; 91-208148/28.

DR Recombinant human factor VIII deriv. deoxyribonucleic acid -

PT encoding protein comprising two chains linked by segment of B domain.

PS Disclosure; Fig 1; 35pp; English.

CC The protein is a fusion between Phe 742 and Ser 1637 of the factor VIII protein (factor VIII:SQ). In order to produce a

CC factor VIII deletion derivative that can be produced in vivo and/or

CC in vitro, to a two chain protein consisting of polypeptide chains of

CC 90 kD and 80 kD, the amino acid sequences surrounding Arg 740 and

CC Arg 1648 have to be conserved in order to preserve the structural

CC requirements for correct cleavage. In this example, amino acids 743

CC to 1636 of the full-length factor VIII polypeptide are deleted. A

CC new polypeptide chain is obt. where there are 14 amino acids

CC linking Arg 740 and Arg 1648. Of these 14 amino acids, the sequence

CC of the five N-terminal ones directly corresponds to the five amino

CC acids following Arg 740 in full-length factor VIII. Also, the sequence

CC of the 12 C-terminal amino acids of the above 14 amino acids fragment

CC directly corresponds to the 12 amino acids preceding Glu 1689 in full-

CC length factor VIII, thus creating a 3 amino acid overlap between the

CC N- and C-terminal regions of the B-domain.

CC The factor VIII deriv. is useful for treating haemophilia or

CC haemophilia A. It has the biological characteristics of plasma derived

CC factor VIII.

CC In order to index this example, the factor VIII:QD amino acid

CC sequence was retrieved from WO8800831 (P80265).

CC The amino acid numbering in the above comments is reproduced from the

CC fig. description in the specification. Note that Arg 740 is Arg 742

CC In the indexed sequence, etc. Also, Asn 745 (N-terminal link overlap)
CC is Asp 747 in P80265, but indexed as Asn to reproduce the fusion
CC fragment as shown in fig.1, and Asn 745 (C-terminal link overlap).
SQ Sequence 1440 AA;

Query Match 98.4%; Score 7668.5; DB 1; Length 1440;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1435; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 20 ATRRYLGLAVELSWDYNQSDLGELPYDARFPPRVKSPFENTSVVYKTLFVEFTDHLFN 79
DB 2 ATRRYLGLAVELSWDYNQSDLGELPYDARFPPRVKSPFENTSVVYKTLFVEFTDHLFN 61

QY 80 IAKPRPVMGLGPTIOAEVYDVTVTLLKNMASHPVSLHAGVSYWKASGAEYDQTSQ 139
DB 62 IAKPRPVMGLGPTIOAEVYDVTVTLLKNMASHPVSLHAGVSYWKASGAEYDQTSQ 121

QY 140 REKEDDKVPPGSHYVWQVLKENGPMASDPLCTYYSLSHVLDLVKDLNSGLIGALLVCR 199
DB 122 REKEDDKVPPGSHYVWQVLKENGPMASDPLCTYYSLSHVLDLVKDLNSGLIGALLVCR 181

QY 200 EGSIAKEKTLHKLFLFAVFEDEGKSWHSETKNSLMQDRDAASARAPKMTVNGVYNR 259
DB 182 EGSIAKEKTLHKLFLFAVFEDEGKSWHSETKNSLMQDRDAASARAPKMTVNGVYNR 241

QY 260 S-LPGLIGCHRSYVWVIGMGTTPPEHSIFLEGHFLVRNHRQASLEISPTFLFAQTL 318
DB 242 SLPLGLIGCHRSYVWVIGMGTTPPEHSIFLEGHFLVRNHRQASLEISPTFLFAQTL 301

QY 319 LMDIGQFLFCHISSHQHDMGEAYVYKVDSCPEEPQLRMKNNEAEYDDDLTDESEMDVR 378
DB 302 LMDIGQFLFCHISSHQHDMGEAYVYKVDSCPEEPQLRMKNNEAEYDDDLTDESEMDVR 361

QY 379 FDDNPSFSFIQIRSYAKKHPTKWVHYIAEEDWDYAPLVLPDDRKYKSOYLNGPQRI 438
DB 362 FDDNPSFSFIQIRSYAKKHPTKWVHYIAEEDWDYAPLVLPDDRKYKSOYLNGPQRI 421

QY 439 GRKYKVRFMAYTDETFKTRIAIOHESGILGPLLYGEVGDTLIIFKNQASRPYNTYPHG 498
DB 422 GRKYKVRFMAYTDETFKTRIAIOHESGILGPLLYGEVGDTLIIFKNQASRPYNTYPHG 481

QY 499 ITDVRPLYSRRLPKGVKHLKDFILPGEIEFKYKWTYVEDGPTKSDPRCLTRYSSFFNM 558
DB 482 ITDVRPLYSRRLPKGVKHLKDFILPGEIEFKYKWTYVEDGPTKSDPRCLTRYSSFFNM 541

QY 559 ERDLASGLIGPLLCYKESVDQRGNOIMSDKRNVLFSVEDENRSWYLENIOQRFLPNPA 618
DB 542 ERDLASGLIGPLLCYKESVDQRGNOIMSDKRNVLFSVEDENRSWYLENIOQRFLPNPA 601

QY 619 GVOLEDPEFOASINMHSINGVYFDSLOLSVCLHEVAYWYILSIGAQIDFLSVFFSGYTFK 678
DB 602 GVOLEDPEFOASINMHSINGVYFDSLOLSVCLHEVAYWYILSIGAQIDFLSVFFSGYTFK 661

QY 679 HKMYVEDTLTLFPFSGETVFMSEMENGLMILGCHNSDFRNRGMTALLKVSCKDKNTGDIY 738
DB 662 HKMYVEDTLTLFPFSGETVFMSEMENGLMILGCHNSDFRNRGMTALLKVSCKDKNTGDIY 721

QY 739 EDSYEDISAYLLSKNNAIEPRFSQNPVLKRHOREITRTTLOSQDEIDYDDTISYEMK 798
DB 722 EDSYEDISAYLLSKNNAIEPRFSQNPVLKRHOREITRTTLOSQDEIDYDDTISYEMK 781

QY 799 KEDFDIYDEDENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHYLRNRAQSGSVPOQPKY 858
DB 782 KEDFDIYDEDENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHYLRNRAQSGSVPOQPKY 841

QY 859 VFQETDGSFTQPLYRGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYE 918
DB 842 VFQETDGSFTQPLYRGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYE 901

QY 919 EDOROGAEPKRNFKVKNETKTYFWKYVQHMAPTKDEFDCKAWAYFSVDVLEKDVHSGLIG 978
DB 902 EDOROGAEPKRNFKVKNETKTYFWKYVQHMAPTKDEFDCKAWAYFSVDVLEKDVHSGLIG 961

QY 979 PLVCHTNTLNPAGROVTVQEFALFTIFDETKSWYFTEEMERNCRAPCNIOEMDPTFK 1038
DB 962 PLVCHTNTLNPAGROVTVQEFALFTIFDETKSWYFTEEMERNCRAPCNIOEMDPTFK 1021

QY 1039 ENYRFAHNGYIMDTPLGLVMAQDQIRWYLLSMGSMENIHSHFSGHVFTVRKKEEYKM 1098
DB 1022 ENYRFAHNGYIMDTPLGLVMAQDQIRWYLLSMGSMENIHSHFSGHVFTVRKKEEYKM 1081

QY 1099 ALYNLYPGVETVEMLPSKAGIWRVECLIGEHLHAGMSTFLVYSNKCOTPLGMASGHIR 1158
DB 1082 ALYNLYPGVETVEMLPSKAGIWRVECLIGEHLHAGMSTFLVYSNKCOTPLGMASGHIR 1141

QY 1159 DFQITASGOYGOMAEKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGAROKF 1218
DB 1142 DFQITASGOYGOMAEKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGAROKF 1201

QY 1219 SSLYISOFTIMYSLDGKKWQTYRGNSTGTLMVFEQVNDSSGKHNIFNPPIARIYRLHP 1278
DB 1202 SSLYISOFTIMYSLDGKKWQTYRGNSTGTLMVFEQVNDSSGKHNIFNPPIARIYRLHP 1261

QY 1279 THYSIRSTLRLMELMGCDLNSCMLPMEKSAISDAQITASSYFTNMFATWSPSKARLHLQ 1338
DB 1262 THYSIRSTLRLMELMGCDLNSCMLPMEKSAISDAQITASSYFTNMFATWSPSKARLHLQ 1321

QY 1339 GRSNMRPQVNNPKEMLOVDFOKTMKVTGVTTOGVKSLTSMYKFEFLISSQDGHQWTL 1398
DB 1322 GRSNMRPQVNNPKEMLOVDFOKTMKVTGVTTOGVKSLTSMYKFEFLISSQDGHQWTL 1381

QY 1399 FFQNGKVKVFOGNQDSFTPVVNSLDPELLTRYLRHPQSWVHQIALRMEVLGCEADQLY 1457
DB 1382 FFQNGKVKVFOGNQDSFTPVVNSLDPELLTRYLRHPQSWVHQIALRMEVLGCEADQLY 1440

RESULT 5

PROT 5
ID P80265 standard; protein; 1516 AA.
AC P80265;
DT 10-OCT-1990 (first entry)
DE Modified factor VIII:C sequence with the Q744-D1563 deletion.
KW Modified factor VIII:C; maturation polypeptide; haemophilia;
KW blood coagulation; QD deletion.
OS Homo sapiens.
PN WO800831-A.
PD 11-FEB-1988.
PF 31-JUL-1987; U01814.
PR 01-AUG-1986; US-893375.
PA (BIOJ) Biogen NV (PASE/).
PI pasek MP;
DR MPI: 88-049866/07.
DR N-PSDB: N80444.
PT New DNA sequences encoding modified factor VIII:C - with deletion of DNA
PT encoding maturation polypeptide, useful for high yield transformation.
PS Claim 3; Page 51-52-53-54; 97pp; English.
CC A major part of the sequence encoding the maturation polypeptide of
CC factor VIII:C is deleted, i.e. Gln 744 - Asp 1563. The QD deletion
CC retains approximately 90 amino acids of the maturation polypeptide
CC (four amino acids at the N-terminal end and 86 amino acids at
CC the C-terminal end). The full length factor VIII:C cDNA has two
CC changes with respect to the published sequence (Epo application 160457):
CC CTG to CTA at Leu 242 and TTC to CTC change at amino acid residue 1880
CC (Phe to Leu).
CC The product is produced in approx. 20 times higher
CC yields than previous recombinant produced factor VIII:C and are more
CC easily purified. The peptide is used for treating haemophilia A, both
CC acute and prolonged bleeding.
CC See also N80446 and N80447.
SQ Sequence 1516 AA;

Query Match 97.7%; Score 7620.5; DB 1; Length 1516;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1435; Conservative 1; Mismatches 2; Indels 77; Gaps 2;

QY 20 ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVPKSPFNTSVYKKTLFVEFTDHLFN 79
Db 2 ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVPKSPFNTSVYKKTLFVEFTDHLFN 61
QY 80 IAKPRPMMGLLPTIQAEVYDVTVITLKNMASHPVSLHAGVSVYWKASGAEYDDQTSQ 139
Db 62 IAKPRPMMGLLPTIQAEVYDVTVITLKNMASHPVSLHAGVSVYWKASGAEYDDQTSQ 121
QY 140 REKEDKVPFGSGSHTYWQVLKENGPMASDPLCLTYSYLSHVLYKDLNSGLIGALLYCR 199
Db 122 REKEDKVPFGSGSHTYWQVLKENGPMASDPLCLTYSYLSHVLYKDLNSGLIGALLYCR 181
QY 200 EGSLAKEKTQTLHKFILLFAVDEGKSMHSETKNSLMQDRDASAARAMPKMTVNGYVNR 259
Db 182 EGSLAKEKTHTLHKFILLFAVDEGKSMHSETKNSLMQDRDASAARAMPKMTVNGYVNR 241
QY 260 S-LPLGLIGHKRSYVWHVIGMGTTPREVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL 318
Db 242 SLLPGLIGHKRSYVWHVIGMGTTPREVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL 301
QY 319 LMDLGQFLFCHISSHODGMEAYVKVDSCBEPOLRMKNNEAEDYDDDLTDSEMDVVR 378
Db 302 LMDLGQFLFCHISSHODGMEAYVKVDSCBEPOLRMKNNEAEDYDDDLTDSEMDVVR 361
QY 379 FDDDNPSFQIRSVAKKHPTWVHYIAAEEDMDYAPLYLAPDDRYSKYQYLNNGPORI 438
Db 362 FDDDNPSFQIRSVAKKHPTWVHYIAAEEDMDYAPLYLAPDDRYSKYQYLNNGPORI 421
QY 439 GRKYKKVRFEMAYTDETFKTREAIQHESGILGPLLYGEVGDILLIFKNQASRPYNIYPHG 498
Db 422 GRKYKKVRFEMAYTDETFKTREAIQHESGILGPLLYGEVGDILLIFKNQASRPYNIYPHG 481
QY 499 ITDVRPLYSRRLPKGVKHLKDFPILPGEIFKXKWTVTVEDGPTKSDPRCLTRYSSFVNM 558
Db 482 ITDVRPLYSRRLPKGVKHLKDFPILPGEIFKXKWTVTVEDGPTKSDPRCLTRYSSFVNM 541
QY 559 ERDLASGLIGPLILICYKESVDQGNQIMSDKRNILFSVFDENRSWYLTENIORFLPNPA 618
Db 542 ERDLASGLIGPLILICYKESVDQGNQIMSDKRNILFSVFDENRSWYLTENIORFLPNPA 601
QY 619 GVOLEDPEFOASNIMHSINGVYVPSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFK 678
Db 602 GVOLEDPEFOASNIMHSINGVYVPSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFK 661
QY 679 HKMYEDTTLTFPFGSETVFMSEMNGLMILGCHNSDFRNRGMTALLKVSSCDKNTGDIY 738
Db 662 HKMYEDTTLTFPFGSETVFMSEMNGLMILGCHNSDFRNRGMTALLKVSSCDKNTGDIY 721
QY 739 EDSYEDISAYLLSKNNAIEPRSF-----SQNPPVLRKHOREITRTLOS 762
Db 722 EDSYEDISAYLLSKNNAIEPRSFQODPLAWDNHYGTQIPKEWKSQEKSPKTAFFKKKDT 781
QY 762 -----SQNPPVLRKHOREITRTLOS 782
Db 782 ILSLNACESNHAIAINEGONKPEIEVTWAKQGRTERLCSQNPVLRKHOREITRTLOS 841
QY 783 DQEEIDYDDTISVEMKKEDEDIYDEDENQSPRSFOKKTRHYFIAAVERLMDYGMSSSPHV 842
Db 842 DQEEIDYDDTISVEMKKEDEDIYDEDENQSPRSFOKKTRHYFIAAVERLMDYGMSSSPHV 901
QY 843 LRNRAGSGSVPOFKKVVFOEFTDGSFTQPLRGELNEHLGLGPIYRAVEDNINMTFRN 902
Db 902 LRNRAGSGSVPOFKKVVFOEFTDGSFTQPLRGELNEHLGLGPIYRAVEDNINMTFRN 961
QY 903 QASRPYSFYSLLSYEEDOROGAEPKRNFKVKNETKTYFWKVQHMAPTKDEFDCKAWAY 962
Db 962 QASRPYSFYSLLSYEEDOROGAEPKRNFKVKNETKTYFWKVQHMAPTKDEFDCKAWAY 1021
QY 963 FSDVDLEKDVHSGLIGPLIVCHTNTLNBAHGRQVTVQEFALFTIIFDETSKSWYFTENNER 1022
Db 1022 FSDVDLEKDVHSGLIGPLIVCHTNTLNBAHGRQVTVQEFALFTIIFDETSKSWYFTENNER 1081

QY 1023 NCRAPCNIQMEDPTFKENYREHAINGYIMDTLLPGLVMAQDQIRIRWYLLSMGSENHISH 1082
Db 1082 NCRAPCNIQMEDPTFKENYREHAINGYIMDTLLPGLVMAQDQIRIRWYLLSMGSENHISH 1141
QY 1083 FSGHVFTRKKEEYKMAVLYLPGVFETVENMLPSKAGIWRVECLLGEHLAGMSTLFLVY 1142
Db 1142 FSGHVFTRKKEEYKMAVLYLPGVFETVENMLPSKAGIWRVECLLGEHLAGMSTLFLVY 1201
QY 1143 SNRCQTPMGASGHIRDFOITASGOYGOWAPKLARLHYSGSINAMSTKEPFWIKVDLLA 1202
Db 1202 SNRCQTPMGASGHIRDFOITASGOYGOWAPKLARLHYSGSINAMSTKEPFWIKVDLLA 1261
QY 1203 PMITHIKTQGARQKFSLSYISQFILMYSLDGKKWQTYRGNSGTLMVFFGNVDSGIKH 1262
Db 1262 PMITHIKTQGARQKFSLSYISQFILMYSLDGKKWQTYRGNSGTLMVFFGNVDSGIKH 1321
QY 1263 NTFNPPITARIYRLHPTHSIRSTLRMELMGCDLNSCMLPGMESKAISDAQITASSYFT 1322
Db 1322 NTFNPPITARIYRLHPTHSIRSTLRMELMGCDLNSCMLPGMESKAISDAQITASSYFT 1381
QY 1323 NMFATWSPSKARLHIOGRSNAMRPQVNNPKEWIYDFOKTMYKTGVTTOGVKSLLTSMYV 1382
Db 1382 NMFATWSPSKARLHIOGRSNAMRPQVNNPKEWIYDFOKTMYKTGVTTOGVKSLLTSMYV 1441
QY 1442 KEFLISSQDGHQWTLFFQNGKVKYFQGNQDSFTPVNSLDPPLTRYLRIRHQSFWHQI 1501
Db 1443 ALRMEVLGCEAODLY 1457
1502 ALRMEVLGCEAODLY 1516

RESULT 6
P80267
ID P80267 standard; protein: 1425 AA.
AC P80267:
DT 10-OCT-1990 (first entry)
DE Modified factor VIII:C sequence with the R740-E1649 deletion.
KW Modified factor VIII:C; maturation polypeptide; haemophilia;
KW blood coagulation; RE deletion.
OS Homo sapiens.
PN W08800831-A.
PD 11-FEB-1988.
PE 31-JUL-1987; U01814.
PR 01-AUG-1986; US-893375.
PA (BioJ) Biogen NV (PASE/).
PI PASEK MP;
DR WPI; 88-049866/07.
DR N-PSDB; N80446.
PT New DNA sequences encoding modified factor VIII:C - with deletion of DNA
PT encoding maturation polypeptide, useful for high yield transformation.
PS Claim 3; Page 57-58-59-60; 97pp; English.
CC The entire sequence encoding the maturation polypeptide of
CC factor VIII:C is deleted, i.e. Arg 740-Glu 1649.
CC The full length Factor VIII:C cDNA has two changes with respect to the
CC published sequence (EPO application 160457):
CC CTG to CTA at Leu 242 and TTC to CTC change at amino acid residue 1880
CC (Phe to Leu). The product is produced in approx. 20 times higher
CC yields than previous recombinant produced factor VIII:C and are more
CC easily purified. The peptide is used for treating haemophilia A, both
CC acute and prolonged bleeding.
CC See also N80444 and N80447.
SQ Sequence 1425 AA;

Query Match 97.3%; Score 7590; DB 1; Length 1425;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1422; Conservative 0; Mismatches 2; Indels 14; Gaps 1;
QY 20 ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVPKSPFNTSVYKKTLFVEFTDHLFN 79
Db 2 ATRRYYLGAVELSWDMQSDLGELPVDARPPRPVPKSPFNTSVYKKTLFVEFTDHLFN 61

Db	181	EGSLAKEKTHTLHKFIILLFAVFEDECKSWHSETKNSLMQDRDAASARAMPKMHVTNGVNR	240
QY	260	SLPGLIGCHRKSVYWHYIGMGTTPREHVSIFLEGHTFLVRNHRQASLEISPTFLTAQTL	319
Db	241	SLPGLIGCHRKSVYWHYIGMGTTPREHVSIFLEGHTFLVRNHRQASLEISPTFLTAQTL	300
QY	320	MDLGQFLLFCHISSHQHDGMEAYVKVDSCEPEPQLRMKNNEAEADYDDDLTDEMDVVR	379
Db	301	MDLGQFLLFCHISSHQHDGMEAYVKVDSCEPEPQLRMKNNEAEADYDDDLTDEMDVVR	360
QY	380	DDNSPSFIQIRSVAKKHPKTWVHYIAAEEDWDYAPLVLAPDDRSYKSQYLNNGPORIG	439
Db	361	DDNSPSFIQIRSVAKKHPKTWVHYIAAEEDWDYAPLVLAPDDRSYKSQYLNNGPORIG	420
QY	440	RKYKVRFMAYTDEFKTRERAIQHESGILGPLLYGEVGDTLIIFKNQASRPYNIYPHGI	499
Db	421	RKYKVRFMAYTDEFKTRERAIQHESGILGPLLYGEVGDTLIIFKNQASRPYNIYPHGI	480
QY	500	TDVRLPLSRRLPKGVKHLKDFPILPGELIFKYKWTVVEDGPTKSDPRCLTRYSSSEFNME	559
Db	481	TDVRLPLSRRLPKGVKHLKDFPILPGELIFKYKWTVVEDGPTKSDPRCLTRYSSSEFNME	540
QY	560	RDLASGLIGPLLCYKESVDQRGNOIMSDKRNVLFSVFDENRSWYLTENIQRFPLNPAG	619
Db	541	RDLASGLIGPLLCYKESVDQRGNOIMSDKRNVLFSVFDENRSWYLTENIQRFPLNPAG	600
QY	620	VQLEDPPEQASNMHSINGYVDSLOLSVCLHEVAYWYILSIGAQDFLSVFFSGYTEFKH	679
Db	601	VQLEDPPEQASNMHSINGYVDSLOLSVCLHEVAYWYILSIGAQDFLSVFFSGYTEFKH	660
QY	680	KMYVEDTLTLPFSGEYFVMSMENPGMLWILGCHNSDFRNGMTALLKYSSCDKNTGDIYE	739
Db	661	KMYVEDTLTLPFSGEYFVMSMENPGMLWILGCHNSDFRNGMTALLKYSSCDKNTGDIYE	720
QY	740	DSYEDISAYLLSKNNAIEPRSFQNPVYLKRHQREITRTLTQSDQEEIDYDDTISVEMKK	799
Db	721	DSYEDISAYLLSKNNAIEP-----REITRTLTQSDQEEIDYDDTISVEMKK	766
QY	800	EDFDIYDEENQSPRSFQKTRHYFIAAVERLMDYGMSSGPHVLRNRAQSGSVPOFKKVV	859
Db	767	EDFDIYDEENQSPRSFQKTRHYFIAAVERLMDYGMSSGPHVLRNRAQSGSVPOFKKVV	826
QY	860	FOEFTDGSFTQPLRGELNEHLGLPYIRAEVEDNIMVTFRNQASRPYSFSSLSIYEE	919
Db	827	FOEFTDGSFTQPLRGELNEHLGLPYIRAEVEDNIMVTFRNQASRPYSFSSLSIYEE	886
QY	920	DORQGAEPKRNFKVKNETKYFWKVQHMAPTKDEFDCAKAWAIFSDVDLEKDVHSGLICP	979
Db	887	DORQGAEPKRNFKVKNETKYFWKVQHMAPTKDEFDCAKAWAIFSDVDLEKDVHSGLICP	946
QY	980	LLVCHTNTLNPAGROVTVQEFALFTIFDETKSWYFTENNERNCRAPCNIQMEDPTFKE	1039
Db	947	LLVCHTNTLNPAGROVTVQEFALFTIFDETKSWYFTENNERNCRAPCNIQMEDPTFKE	1006
QY	1040	NYRFHAINGYIMDTLPGLVMAQDORIRWYLLSMGSNENIHSHFSGHVFTYRKKEEYKMA	1099
Db	1007	NYRFHAINGYIMDTLPGLVMAQDORIRWYLLSMGSNENIHSHFSGHVFTYRKKEEYKMA	1066
QY	1100	LYNLPGVFEVTEMLPSKAGIWEVECLIGEHLHAGMSTLFLVYSNKCQOTPLGMA SGHIRD	1159
Db	1067	LYNLPGVFEVTEMLPSKAGIWEVECLIGEHLHAGMSTLFLVYSNKCQOTPLGMA SGHIRD	1126
QY	1160	FOITASGOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLAPMIIGIKTOGAROKFS	1219
Db	1127	FOITASGOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLAPMIIGIKTOGAROKFS	1186
QY	1220	SLYISQFIIMYSIDGKKWQYRGNSTGTLMVFFGNVDSGIKHNIENPPIIARIYIRLHPT	1279
Db	1187	SLYISQFIIMYSIDGKKWQYRGNSTGTLMVFFGNVDSGIKHNIENPPIIARIYIRLHPT	1246
QY	1280	HYSIRSTLRMELMGCDLNSCSPMLGEMSKAISDAQITASSYFTNMFPATWSPSKARLHLQG	1339
Db	1247	HYSIRSTLRMELMGCDLNSCSPMLGEMSKAISDAQITASSYFTNMFPATWSPSKARLHLQG	1306

QY	1340	RSNAWRPQVNNPKEWLQVDFOKTMKVGVTTQGVKSLTSMYVKEFLISSQDGHQWTLF	1399
Db	1307	RSNAWRPQVNNPKEWLQVDFOKTMKVGVTTQGVKSLTSMYVKEFLISSQDGHQWTLF	1366
QY	1400	FQNGKVKVFQGNQDSFTPYVNSLDPPLITRYLRIRHPQSVWHQIALRMEVLGCEAQDLY	1457
Db	1367	FQNGKVKVFQGNQDSFTPYVNSLDPPLITRYLRIRHPQSVWHQIALRMEVLGCEAQDLY	1424
RESULT 8			
ID	P91169	standard; protein; 1424 AA.	
AC	P91169;		
DT	26-JUN-1990	(first entry)	
DE	Sequence of 740 Arg-1649 Glu human Factor VIII:C		
KW	Human Factor VIII:C; Ad.RE.neo; 740 Arg-1649 Glu Factor VIII:C;		
KW	haemophililia A.		
OS	Homo sapiens.		
PN	EP-306968-A.		
PD	15-MAR-1989.		
PF	09-SEP-1988; 114769.		
PR	08-APR-1988; JP-085454.		
PA	(KAGA) Chemo-Sero-Therap (Teij).		
PI	Sugiyama T, Masuda K, Tajima Y, Yonemura H;		
DR	WPI; 89-078467/11.		
DR	N-PSDB; N90654.		
PT	Prodn. of recombinant human Factor-VIII-C -		
PT	using animal cells transformed with a vector contg. the gene for		
PT	Factor VIII:C and a promoter		
PS	Figure 1(1) - 1(13); ; 32pp; English.		
CC	Arg-740 of the carboxyl terminus of the H chain is		
CC	directly bonded by a peptide bond to Glu-1649 of the amino terminus of		
CC	l chain. A pref. expression vector used to transform animal cell so		
CC	that they produce human Factor VIII:Cis plasmid Ad.RE.neo.		
CC	The expression vector has at least one promoter upstream of n90654.		
CC	The transformants can constantly and continuously produce human Factor		
CC	VIII:C in high yield on a commercial scale. The human Factor VIII:C so		
CC	produced is considered to corresp. to the smallest species of active and		
CC	intact Factor VIII:C molecules in the human blood plasma. It is useful		
CC	for treating haemophililia A patients.		
SO	Sequence 1424 AA;		
Query Match 97.3%; Score 7583; DB 1; Length 1424;			
Best Local Similarity 98.8%; Pred. No. 0;			
Matches 1421; Conservative 0; Mismatches 3; Indels 14; Gaps 1;			
QY	20	ATRRYLLGAVELSMWYQSDLGELPVDARFPFRYPKSPFNTSVYKTLFVEFTDHLFN	79
Db	1	ATRRYLLGAVELSMWYQSDLGELPVDARFPFRYPKSPFNTSVYKTLFVEFTDHLFN	60
QY	80	IAKPRPWWGLLPRTIQAEVYDVTIVITLKNMASHPYSLHAGVSYWKASEGAEXDDQTSQ	139
Db	61	IAKPRPWWGLLPRTIQAEVYDVTIVITLKNMASHPYSLHAGVSYWKASEGAEXDDQTSQ	120
QY	140	REKEDDKVFPGGSHTYVQVLKENGPMASDPLCLTYSYLSHVLYKDLNSGLIGALLVCR	199
Db	121	REKEDDKVFPGGSHTYVQVLKENGPMASDPLCLTYSYLSHVLYKDLNSGLIGALLVCR	180
QY	200	EGSLAKEKTQTLHKFIILLFAVFEDECKSWHSETKNSLMQDRDAASARAMPKMHVTNGVNR	259
Db	181	EGSLAKEKTQTLHKFIILLFAVFEDECKSWHSETKNSLMQDRDAASARAMPKMHVTNGVNR	240
QY	260	SLPGLIGCHRKSVYWHYIGMGTTPREHVSIFLEGHTFLVRNHRQASLEISPTFLTAQTL	319
Db	241	SLPGLIGCHRKSVYWHYIGMGTTPREHVSIFLEGHTFLVRNHRQASLEISPTFLTAQTL	300
QY	320	MDLGQFLLFCHISSHQHDGMEAYVKVDSCEPEPQLRMKNNEAEADYDDDLTDEMDVVR	379
Db	301	MDLGQFLLFCHISSHQHDGMEAYVKVDSCEPEPQLRMKNNEAEADYDDDLTDEMDVVR	360
QY	380	DDNSPSFIQIRSVAKKHPKTWVHYIAAEEDWDYAPLVLAPDDRSYKSQYLNNGPORIG	439

Db 361 DDDNSPSFQIRSVAKKHPTWVHYIAAEEDWDYAPLYLAPHDRSYKSOYLNNGPORIG 420
QY 440 RKKKKVRFMAVTDETFKTREAIQIHESGILGPLLYGEVGDILLIFKNQASRPYNIYPHGI 499
Db 421 RKKKKVRFMAVTDETFKTREAIQIHESGILGPLLYGEVGDILLIFKNQASRPYNIYPHGI 480
QY 500 TDVRPLYSRRLPKGVKHLKDPILPGEIFKXKWTVTVEDPPTKSDPRCLTRYSSFYVME 559
Db 481 TDVRPLYSRRLPKGVKHLKDPILPGEIFKXKWTVTVEDPPTKSDPRCLTRYSSFYVME 540
QY 560 RDLASGLIGPLLICKESVDQRGNOIMSDKRNVLFSVFEDENRSWYLTENIQRFLLPNPAG 619
Db 541 RDLASGLIGPLLICKESVDQRGNOIMSDKRNVLFSVFEDENRSWYLTENIQRFLLPNPAG 600
QY 620 VQLEDPFQASNIMHSINGVFDLSQLSVCLHEVAYWYIISGAQTDFLSVFFSGYTFKH 679
Db 601 VQLEDPFQASNIMHSINGVFDLSQLSVCLHEVAYWYIISGAQTDFLSVFFSGYTFKH 660
QY 680 KMYVEDTLTLFPFSGEIVMSMENPGLWITGCHNSDFRNKGMTALLKVSSCDKNTGDYYE 739
Db 661 KMYVEDTLTLFPFSGEIVMSMENPGLWITGCHNSDFRNKGMTALLKVSSCDKNTGDYYE 720
QY 740 DSYEDISAYLLSKNNAIEPSPFSQNPVULKHOREITRTTLQSDQEEIDYDDTISVEMKK 799
Db 721 DSYEDISAYLLSKNNAIEP-----REITRTTLQSDQEEIDYDDTISVEMKK 766
QY 800 EDFDIYDEENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHYLRNRAQSGSVPOFKKVV 859
Db 767 EDFDIYDEENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHYLRNRAQSGSVPOFKKVV 826
QY 860 FOEFTDGSFTQPLRYRGELNHLGLGPYIRAEVEDNIMWTFERNQASRPYSFYSLSISYEE 919
Db 827 FOEFTDGSFTQPLRYRGELNHLGLGPYIRAEVEDNIMWTFERNQASRPYSFYSLSISYEE 886
QY 920 DOROGAEPKKNFVKPNETKTYFWKVQHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGP 979
Db 887 DOROGAEPKKNFVKPNETKTYFWKVQHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGP 946
QY 980 LLVCHTNTLNPAGROVTVQVEFALFTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKE 1039
Db 947 LLVCHTNTLNPAGROVTVQVEFALFTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKE 1006
QY 1040 NYRFHAINGYIMDTLPGLVNAQDORIRWYLYLSMGSENENIHSHFSGHVFTVRKKEEYKMA 1099
Db 1007 NYRFHAINGYIMDTLPGLVNAQDORIRWYLYLSMGSENENIHSHFSGHVFTVRKKEEYKMA 1066
QY 1100 LYNLYPGVEETVEMLPKSKAGIWRVECLIGELHLAGMSTLFLVYSNKCQTPGLMASGHIRD 1159
Db 1067 LYNLYPGVEETVEMLPKSKAGIWRVECLIGELHLAGMSTLFLVYSNKCQTPGLMASGHIRD 1126
QY 1160 FOITASGOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGICTOGAROKFS 1219
Db 1127 FOITASGOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGICTOGAROKFS 1186
QY 1220 SLYISQFIIMYSLDGKKWQYTRGNSTGTLMWYFEGNVDDSGIKHNIENPPIIARYIRLHPT 1279
Db 1187 SLYISQFIIMYSLDGKKWQYTRGNSTGTLMWYFEGNVDDSGIKHNIENPPIIARYIRLHPT 1246
QY 1280 HYSIRSTLRMELMGCDLNSCMLPGMESKATSDAQITASSYFTNMFAWSPSKARLHLQG 1339
Db 1247 HYSIRSTLRMELMGCDLNSCMLPGMESKATSDAQITASSYFTNMFAWSPSKARLHLQG 1306
QY 1340 RSNAMRPQVNNPKEMLOVDFOKTMKVTVGTVOGVKSLTSMYVKEFLISSSDGHQWTLF 1399
Db 1307 RSNAMRPQVNNPKEMLOVDFOKTMKVTVGTVOGVKSLTSMYVKEFLISSSDGHQWTLF 1366
QY 1400 FONGKVKVFQGNQDSTPVVNSLDPLTLTRYLRIHQSWVHQIALRMEVLCGEAQDLY 1457
Db 1367 FONGKVKVFQGNQDSTPVVNSLDPLTLTRYLRIHQSWVHQIALRMEVLCGEAQDLY 1424

RESULT 9

W18670
ID W18670 standard; Protein; 1661 AA.
AC W18670;
DE 10-AUG-1997 (first entry)
KW Factor VIII-dB695-HCII.
KW Factor VIII-dB695-HCII; heparin cofactor II; blood coagulation;
KW blood clotting; procoagulant; anticoagulant; antithrombotic;
KW haemophilia; gene therapy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT domain 731..760
FT /label= HCII
FT /note= "heparin cofactor II acidic domain"
PI W09718315-A1.
DR 22-MAY-1997.
PD 13-NOV-1996; E04977.
PR 13-NOV-1995; US-558107.
PA (IMMO) IMMUNO AG.
PI Voorberg Jf;
DR WPI; 97-289291/26.
DR N-PSDB; T69811.
PT Hybrid Factor VIII with modified activity, comprises region from
PT donor anticoagulant or antithrombotic protein - useful for treatment
PT of coagulation disorders
PS Claim 11; Page 52-60; 96pp; English.
CC Factor VIII-dB695-HCII (W18670) is a hybrid protein in which amino
CC acids 712-736 of Factor-dB695 (Factor VIII del868-1562) B-domain
CC are replaced by amino acids 51-80 from the acidic region (and
CC potential thrombin-binding site) of human heparin cofactor II
CC (HCII). It is the expression product of Factor VIII-dB695-HCII
CC DNA (T69811) in plasmid pCLB-dB695-HCII. The hybrid protein, which
CC can be expressed using gene therapy techniques, has increased
CC procoagulant activity owing to the HCII acidic region, and can be
CC used to treat blood coagulation disorders such as haemophilia A.
SQ Sequence 1661 AA;

Query Match 96.8%; Score 7544; DB 1; Length 1661;
Best Local Similarity 86.2%; Pred. No. 0;
Matches 1437; Conservative 5; Mismatches 9; Indels 216; Gaps 4;
QY 1 MQIELSTCFEFLCLLRFCSATRRYYLAGVELSWDMQSDGELPVDARPPRPVPSFPFN 60
Db 1 MEIELSTCFEFLCLLRFCSATRRYYLAGVELSWDMQSDGELPVDARPPRPVPSFPFN 60
QY 61 TSVVYKKTLEVEFTDHLFNIAKPRPMMGLGPTIQAEVYDVTVITLKNASHPVSLHAY 120
Db 61 TSVVYKKTLEVEFTDHLFNIAKPRPMMGLGPTIQAEVYDVTVITLKNASHPVSLHAY 120
QY 121 GVSYWKAEGAEYDDQTSOREKEDDKVFPGSGHTYVWQYLKENGPMASDPLCLTYSYLSH 180
Db 121 GVSYWKAEGAEYDDQTSOREKEDDKVFPGSGHTYVWQYLKENGPMASDPLCLTYSYLSH 180
QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRD 240
Db 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRD 240
QY 241 AASARAWPMHTVNGVYVNSLPLGLICGCHKSYYWHVIGMGTTPREVHSIFLEGHTFLVRNH 300
Db 241 AASARAWPMHTVNGVYVNSLPLGLICGCHKSYYWHVIGMGTTPREVHSIFLEGHTFLVRNH 300
QY 301 RQASLEISPIFTLTAQTLMDLGQFLLECHIISSHQHDCMEAYVKVDSCEEPQOLRMKNE 360
Db 301 RQASLEISPIFTLTAQTLMDLGQFLLECHIISSHQHDCMEAYVKVDSCEEPQOLRMKNE 360
QY 361 EAEDYDDDLTDSMDVVRDDDDNSPSFQIRSVAKKHPTWVHYIAAEEDWDYAPLYLA 420
Db 361 EAEDYDDDLTDSMDVVRDDDDNSPSFQIRSVAKKHPTWVHYIAAEEDWDYAPLYLA 420
QY 421 PDDRSYKSOYLNNGPORIGRKKYKVRFMAVTDETFKTREAIQIHESGILGPLLYGEVGDIL 480
Db 421 PDDRSYKSOYLNNGPORIGRKKYKVRFMAVTDETFKTREAIQIHESGILGPLLYGEVGDIL 480

VIII, along with calcium and phospholipid, acts as a cofactor for factor IXA, when it converts factor X to the activated form (factor XA). FVIII is the coagulation factor deficient in the X-chromosome-linked bleeding disorder haemophilia A. Several other mutant FVIII proteins have also been created (see W33222-29). The FVIII mutant F309S (W33225) is capable of recombinant secretion at higher levels than typically obtained with wild type FVIII and retains pro-coagulant activity. The FVIII mutant R336I (W33222) and R562K (W33223) are resistant to activated protein C (APC) cleavage. The present FVIII mutant can form a more stable configuration, and have an approximate 5-fold increase in specific activity compared to purified wild type FVIII, while increasing their binding affinity to von Willebrand factor improves their stability. The FVIII proteins can be administered to haemophiliacs, i.e. FVIII replacement therapy, while the nucleic acid molecule can be used for gene therapy.

note: this sequence does not appear in the specification; it was created using sequences from the given references.

Sequence 1383 AA:

Query Match	94.4%;	Score 7362.5;	DB 1;	Length 1383;
Best Local Similarity	96.1%;	Pred. No. 0;		
Matches 1382; Conservative	0;	Mismatches	1;	Indels 55; Gaps 14;

QY	20	ATRRYYLGAVELSWDMQSDLGELPYDARFPRVPKSPENTSVYKKTLFVEFTDHLFN	79
Db	1	ATRRYYLGAVELSWDMQSDLGELPYDARFPRVPKSPENTSVYKKTLFVEFTDHLFN	60
QY	80	IAKPRPPWMLGPTIOAEVYDVTVTLLKNMASHPVSLHAGVSYWKASEGAEDDQTSQ	139
Db	61	IAKPRPPWMLGPTIOAEVYDVTVTLLKNMASHPVSLHAGVSYWKASEGAEDDQTSQ	120
QY	140	REKEDDKVFPGGSHYYVQVLKENGPMASDPLCLTYSLSHVDLYKDINSGLIGALLYCR	199
Db	121	REKEDDKVFPGGSHYYVQVLKENGPMASDPLCLTYSLSHVDLYKDINSGLIGALLYCR	180
QY	200	EGSLAKEKTQTLHKFTLLFAVDEGKSMHSETKNSLMODRDAASARAWPMHTVNGYVR	259
Db	181	EGSLAKEKTQTLHKFTLLFAVDEGKSMHSETKNSLMODRDAASARAWPMHTVNGYVR	240
QY	260	SLPGLIGCHRSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLESPITFLTAQTL	319
Db	241	SLPGLIGCHRSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLESPITFLTAQTL	300
QY	320	MDLGOFLLFCHISSHQHDMGEAYKVYDSCPEEPQLRMKNEEAEDYDDDLTDESEMDVVR	379
Db	301	MDLGOFLLFCHISSHQHDMGEAYKVYDSCPEEPQLRMKNEEAEDYDDDLTDESEMDVVR	360
QY	380	DDNDSPSFIQIRSVAKKHPTWVHYIAEEDMDYAPLYLAPDDRYSKQYLNNGPQRIG	439
Db	361	DDNDSPSFIQIRSVAKKHPTWVHYIAEEDMDYAPLYLAPDDRYSKQYLNNGPQRIG	420
QY	440	RKYKVRFMAYTDETFKTRERAIQHESGILGPLLYGEVGDTLIIFKNOASRPYNIPHGI	499
Db	421	RKYKVRFMAYTDETFKTRERAIQHESGILGPLLYGEVGDTLIIFKNOASRPYNIPHGI	480
QY	500	TDVRPLYSRRLPKGVKHLKDFILPGELEFYKWTVTVEDEGPTKSDPRCLTRYSSFVME	559
Db	481	TDVRPLYSRRLPKGVKHLKDFILPGELEFYKWTVTVEDEGPTKSDPRCLTRYSSFVME	540
QY	560	RDLASGLIGPLLCIKYESVDORGNOIMSDKRNVLFSVFEDENRSWYLTENIQRFPLNPAG	619
Db	541	RDLASGLIGPLLCIKYESVDORGNOIMSDKRNVLFSVFEDENRSWYLTENIQRFPLNPAG	600
QY	620	VQLEDEPFOASNIMHSINGYVFDLSQLSVCLHEVAYWYLLSIGAQTDPLSVFFSGYTFKH	679
Db	601	VQLEDEPFOASNIMHSINGYVFDLSQLSVCLHEVAYWYLLSIGAQTDPLSVFFSGYTFKH	660
QY	680	KMYEDTLTLFPFSGEYVFMSEMPGLMTILGCHNSDFRNRGMTALLKVSSCDKNTGDYYE	739
Db	661	KMYEDTLTLFPFSGEYVFMSEMPGLMTILGCHNSDFRNRGMTALLKVSSCDKNTGDYYE	720
QY	740	DSYEDISAYLLSKNNAIEPRFSQONPVLKRHOREITRTTLQSDQOEIDYDQTSIVEMRK	799

Dd	721	DSYEDISAYLLSKNAIE-----	739
QY	800	EDFDIYDEDENOSPRSFOKTRHYFAAVERLMDYGMSSSPVLRNRAOGSGVPQFKKV	859
Dd	739	-----PASFOKTRHYFAAVERLMDYGMSSSPVLRNRAOGSGVPQFKKV	785
QY	860	FOEFTDGSFTQPLRYGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEE	919
Dd	786	FOEFTDGSFTQPLRYGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEE	845
QY	920	DOROGAEPKRNFKVAPNETKTYFKVQVHHMAPTKDEFDCKAMAYFSDVDLEKDVHSLGIP	979
Dd	846	DOROGAEPKRNFKVAPNETKTYFKVQVHHMAPTKDEFDCKAMAYFSDVDLEKDVHSLGIP	905
QY	980	LLVCHTNTLNPAGHQVTVQOEFALEFTIFDETKSWYFTENMERNCRAPCNIOMEDPTKE	1039
Dd	906	LLVCHTNTLNPAGHQVTVQOEFALEFTIFDETKSWYFTENMERNCRAPCNIOMEDPTKE	965
QY	1040	NYRFAHNGYIMDTPLPGLVMAQDORIRWYLLSGNSNENIHSIHFSGHVFTYAKKEEYKMA	1099
Dd	966	NYRFAHNGYIMDTPLPGLVMAQDORIRWYLLSGNSNENIHSIHFSGHVFTYAKKEEYKMA	1025
QY	1100	LYNLXPVGFETVEMLPSKAGIWRVECLIGEHLHAGMSTLFVYSNKCQOTPLGMA SGHIRD	1159
Dd	1026	LYNLXPVGFETVEMLPSKAGIWRVECLIGEHLHAGMSTLFVYSNKCQOTPLGMA SGHIRD	1085
QY	1160	FOITASGOYGOWAPKLARLHSGSINAWSTKEPFSWIKVDLAPMIHGIKTOGAROKFS	1219
Dd	1086	FOITASGOYGOWAPKLARLHSGSINAWSTKEPFSWIKVDLAPMIHGIKTOGAROKFS	1145
QY	1220	SLYISOFIIMYSLDGKKWQTYRGNSTGTLMEFFGNVDSGIGIHNIENPPIIARYIRLHPT	1279
Dd	1146	SLYISOFIIMYSLDGKKWQTYRGNSTGTLMEFFGNVDSGIGIHNIENPPIIARYIRLHPT	1205
QY	1280	HSISITRLMELMGCDLNSCSMPLGMESKAISDAQITASSYITNMFATWSPSKARLHLOG	1339
Dd	1206	HSISITRLMELMGCDLNSCSMPLGMESKAISDAQITASSYITNMFATWSPSKARLHLOG	1265
QY	1340	RSNAMPQOVNPNPKEMLOVDFOKTMKVTGVTTOGVKSLLTSMYKEFLISSODGHQWTLF	1399
Dd	1266	RSNAMPQOVNPNPKEMLOVDFOKTMKVTGVTTOGVKSLLTSMYKEFLISSODGHQWTLF	1325
QY	1400	FONGKXVVFQGNODSFTPVVNSLDPPLTRYLRIRHPQSWHQTALRMEVLGCEAODLY	1457
Dd	1326	FONGKXVVFQGNODSFTPVVNSLDPPLTRYLRIRHPQSWHQTALRMEVLGCEAODLY	1383

RESULT	11
ID	W333229 standard; protein; 1383 AA.
AC	W333229;
DT	30-APR-1998 (first entry)
DE	Procoagulant-active human factor VIII:C (FVIII) mutant protein
KW	pro-coagulant active factor VIII; FVIII; haemophilia A;
KW	recombinant secretion; pro-coagulant activity; resistance;
KW	activated protein C cleavage; APC; B domain; A2 domain; A3 domain
KW	von Willebrand factor binding site; binding affinity;
KW	FVIII replacement therapy.
OS	Synthetic.
OS	Homo sapiens.
FH	Key
FT	Region
FT	/note=.346
FT	/note="factor VIIIA heavy chain"
FT	741..1383
FT	/note="factor VIIIA light chain"
FT	1..329
FT	/note="A1 domain"
FT	1..179
FT	/note="plastocyanin-like domain 1"
FT	187..329
FT	/note="plastocyanin-like domain 2"
FT	380..711
FT	Domain

Qy	1220	SLVISQFIIMYSLDGKKWQTYRGNSTGLTWFEFGNVDSGIRKNIENPPIARIYRLHPT	1279
Db	1146	SLVISQFIIMYSLDGKKWQTYRGNSTGLTWFEFGNVDSGIRKNIENPPIARIYRLHPT	1205
Qy	1280	HSIRSTLRMELMGCDLNSCSMPLGESKAISDAQITASSYFTNMFATWSPSKARLHLQG	1339
Db	1206	HSIRSTLRMELMGCDLNSCSMPLGESKAISDAQITASSYFTNMFATWSPSKARLHLQG	1265
Qy	1340	RSNAWRPQVNNPREWLQVDFQRTMKVTVGTQGVKSLLTSMYKEFLISSQDGHQWTLF	1399
Db	1266	RSNAWRPQVNNPREWLQVDFQRTMKVTVGTQGVKSLLTSMYKEFLISSQDGHQWTLF	1325
Qy	1400	FQNGKVKVFQGNDSFTPVVNSLDPEPLLTRYLRIHPQSWHQIALRMEVLGCEAODLY	1457
Db	1326	FQNGKVKVFQGNDSFTPVVNSLDPEPLLTRYLRIHPQSWHQIALRMEVLGCEAODLY	1383
RESULT 12			
W33228			
W33228 standard; protein; 1383 AA.			
W33228:			
30-APR-1998 (first entry)			
DE	Procoagulant-active human factor VIIIC (FVII) mutant protein.		
KW	Pro-coagulant active factor VIII: FVIII: haemophilia A;		
KW	recombinant secretion; pro-coagulant activity; resistance;		
KW	activated protein C cleavage; APC; B domain; A2 domain; A3 domain;		
KW	von Willebrand factor binding site; binding affinity;		
KW	FVIII replacement therapy.		
OS	Synthetic.		
OS	Homo sapiens.		
FH	Key	Location/Qualifiers	
FT	Region	1..346	
FT		/note="factor VIIIA heavy chain"	
FT	Region	741..1383	
FT		/note="factor VIIIA light chain"	
FT	Domain	1..329	
FT		/note="A1 domain"	
FT	Domain	1..179	
FT		/note="plastocyanin-like domain 1"	
FT	Domain	187..329	
FT		/note="plastocyanin-like domain 2"	
FT	Domain	380..711	
FT		/note="A2 domain"	
FT	Misc_feature	711..746	
FT		/note="a spacer of the sequence	
FT		SFSQNSRHPSTROKOFNATTPENDIEKTDPMF	
FT		AHRTMPKIQNVSSDLMLL is inserted	
FT		between domains A2 and A3"	
FT	Domain	380..554	
FT		/note="plastocyanin-like domain 3"	
FT	Domain	564..711	
FT		/note="plastocyanin-like domain 4"	
FT	Domain	746..1073	
FT		/note="A3 domain"	
FT	Domain	1073..1221	
FT		/note="C1 domain"	
FT	Domain	1226..1378	
FT		/note="C2 domain"	
FT	Cleavage_site	372..373	
FT		/note="by thrombin"	
FT	Disulfide_bond	153..179	
FT		/note="probable"	
FT	Disulfide_bond	528..554	
FT		/note="probable"	
FT	Misc_difference	336	
FT		/label="R336I	
FT	Misc_difference	562	
FT		/label="R562K	
FT	Misc_difference	740	
FT		/label="R740A	

FT /note= "wild type Arg replaced with Ala"

PN MO9740145-A1.

PD 30-OCT-1997.

PF 24-APR-1997; U06563.

PR 15-MAY-1996; US-017785.

PR 24-APR-1996; US-016117.

PA (UNMI) UNIV MICHIGAN.

PI Amano K, Kaufman RJ, Pipe SW;

DR WPI; 97-535830/49.

PT Modified human pro-coagulant active factor VIII - can be

PT administered to haemophiliacs, i.e. factor VIII replacement therapy

PS Claim 18; Page -; 57pp; English.

CC The present sequence represents a novel pro-coagulant active factor

CC VIII (FVIII) mutant protein, comprising a deletion of the B domain and

CC von Willebrand factor binding site, mutations R336I, R562K and R740A and

CC an addition of an amino acid sequence spacer between the A2 and A3

CC domains. Factor VIII, along with calcium and phospholipid, acts as a

CC cofactor for factor IXa, when it converts factor X to the activated form

CC (factor Xa). FVIII is the coagulation haemophilia A. Several other mutant

CC X-chromosome-linked bleeding disorder haemophilia A. Several other mutant

CC FVIII proteins have also been created (see W33222-29). The FVIII mutant

CC F309S (W33225) is capable of recombinant secretion at higher levels than

CC typically obtained with wild type FVIII and retains pro-coagulant

CC activity. The FVIII mutant R336I (W33222) and R562K (W33223) are

CC resistant to activated protein C (APC) cleavage. The FVIII mutant

CC comprising a deletion of the B domain and von Willebrand factor

CC binding site, a mutation at Arg740 and an addition of an amino acid

CC sequence spacer between the A2 and A3 domains can form a more stable

CC configuration, and have an approximate 5-fold increase in specific

CC activity compared to purified wild type FVIII, while increasing their

CC binding affinity to von Willebrand factor improves their stability.

CC The FVIII proteins can be administered to haemophiliacs, i.e. FVIII

CC replacement therapy, while the nucleic acid molecule can be used for

CC gene therapy.

CC note: this sequence does not appear in the specification; it was created

CC using sequences from the given references.

CC Sequence 1383 AA;

SQ

Query Match 94.3%; Score 7351.5; DB 1; Length 1383;

Best Local Similarity 96.0%; Pred. No. 0;

Matches 1380; Conservative 1; Mismatches 2; Indels 55; Gaps 1;

QY 20 ATRRYVLGAVELSDWYQMSDGLGELPYDARFPPRVKSPFNTSVYKKTLFVEFTDHLFN 79

DB 1 ATRRYVLGAVELSDWYQMSDGLGELPYDARFPPRVKSPFNTSVYKKTLFVEFTDHLFN 60

QY 80 IAKPRPVMGLGPTIOAEVYDVTVTTLKNMASHPVS LHAVGVS YWKA SEGA EYDQTSQ 139

DB 61 IAKPRPVMGLGPTIOAEVYDVTVTTLKNMASHPVS LHAVGVS YWKA SEGA EYDQTSQ 120

QY 140 REKEDKVPFGGSHYVQVLKENGPMASDPLCLTYSLSHVDLYKDLNSGLIGALLVCR 199

DB 121 REKEDKVPFGGSHYVQVLKENGPMASDPLCLTYSLSHVDLYKDLNSGLIGALLVCR 180

QY 200 EGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 259

DB 181 EGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR 240

QY 260 SLPLGLIGCHRSVYWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTITLTAQTL 319

DB 241 SLPLGLIGCHRSVYWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPTITLTAQTL 300

QY 320 MDLGQFLFCHISSHQHDGMEAYVKYDSCPEEPQLRMKNNEAEYDDDLTDESEMDVVR 379

DB 301 MDLGQFLFCHISSHQHDGMEAYVKYDSCPEEPQLRMKNNEAEYDDDLTDESEMDVVR 360

QY 380 DDDNSPSFIQIRSVAKKPKTWVHYIAAEEDWDYAPLV LAPDDRYSKSOYLNNGBORIG 439

DB 361 DDDNSPSFIQIRSVAKKPKTWVHYIAAEEDWDYAPLV LAPDDRYSKSOYLNNGBORIG 420

QY 440 RYKRYVEMAYTDEFTKTRALIQHESGILGPLLYGEVGDTLIIIFKNQASRPYNIYPHGI 499

Db 421 RYKKVRFMAYTDETEKTRERAIQHESGILGPLLYGEVDTLLIFRNQASRPYNTYPHGI 480
QY 500 TDVRFPLYSRRLPKGYKHLKDFPILPGEIFKYYKWYTVEDGPTKSDPRCLTRYSSPVNME 559
Db 481 TDVRFPLYSRRLPKGYKHLKDFPILPGEIFKYYKWYTVEDGPTKSDPRCLTRYSSPVNME 540
QY 560 RDLASGLIGPLLICYESVDQRGNQMSDKRNVLFSVFDENRSWLTENTIQRFPLNPAG 619
Db 541 RDLASGLIGPLLICYESVDQRGNQMSDKRNVLFSVFDENRSWLTENTIQRFPLNPAG 600
QY 620 VQLEDPFQASNMHSINGVYFDSLSVCLHEVAYWYILSIGAQTDFLSVFSGYTFKH 679
Db 601 VQLEDPFQASNMHSINGVYFDSLSVCLHEVAYWYILSIGAQTDFLSVFSGYTFKH 660
QY 680 KMYEDTLTLFPFSGETVFMSENGMLILGCHNSDFRNRMGTALIKVSSCDKNTGDIYE 739
Db 661 KMYEDTLTLFPFSGETVFMSENGMLILGCHNSDFRNRMGTALIKVSSCDKNTGDIYE 720
QY 740 DSYEDISAYLLSKNNAIEPRSFQSDNPVLKRHQRETRTTLOSDEEIDYDDTISVEMKK 799
Db 721 DSYEDISAYLLSKNNAIE----- 739
QY 800 EDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRNQAQSGSVPOEKVV 859
Db 739 -----PASFOKTRHYFIAAVERLWDYGMSSSPHVLNRNQAQSGSVPOEKVV 785
QY 860 FQFTDGSFTQPLRGELNEHLGLCPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEE 919
Db 786 FQFTDGSFTQPLRGELNEHLGLCPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEE 845
QY 920 DQROGAEPKKNFVKNETKTYFWKYQHMAPTKDEFDCKAWAYFSDVLEKDVHSGILGP 979
Db 846 DQROGAEPKKNFVKNETKTYFWKYQHMAPTKDEFDCKAWAYFSDVLEKDVHSGILGP 905
QY 980 LLYCHTNTLNPAHGROYTVQEFALFTTIFDETKSWYFTENMERNCAPCNIQMEDPTKE 1039
Db 906 LLYCHTNTLNPAHGROYTVQEFALFTTIFDETKSWYFTENMERNCAPCNIQMEDPTKE 965
QY 1040 NYRHAINGYIMDTPLGLVMAQDQIRRWYLLSMGSMENIHSHESGHVFTVRKKEEYKMA 1099
Db 966 NYRHAINGYIMDTPLGLVMAQDQIRRWYLLSMGSMENIHSHESGHVFTVRKKEEYKMA 1025
QY 1100 LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKCQTPGLMASGHIRD 1159
Db 1026 LYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKCQTPGLMASGHIRD 1085
QY 1160 FQITASGQYGQWAPKLARLHYSGSINAMSTKEPESWIKVDLAPMIHGIKTOGAROKFS 1219
Db 1086 FQITASGQYGQWAPKLARLHYSGSINAMSTKEPESWIKVDLAPMIHGIKTOGAROKFS 1145
QY 1220 SLVYSQFIIMYSLDGKKQWYTRGNSTGTLMVFFGNVDSSGICKHNINPPIARIYRLHPT 1279
Db 1146 SLVYSQFIIMYSLDGKKQWYTRGNSTGTLMVFFGNVDSSGICKHNINPPIARIYRLHPT 1205
QY 1280 HYSIRSTLRMEIMGCDLNSCSMPLGNEKAISDAQITASSYFTNMFATWSPSKARHLQG 1339
Db 1206 HYSIRSTLRMEIMGCDLNSCSMPLGNEKAISDAQITASSYFTNMFATWSPSKARHLQG 1265
QY 1340 RSNAMRPQVNNPKEMLOVDQKIMKYTVTTQGVKSLTSMYVKEFLISSQDGHQWTLF 1399
Db 1266 RSNAMRPQVNNPKEMLOVDQKIMKYTVTTQGVKSLTSMYVKEFLISSQDGHQWTLF 1325
QY 1400 FQNGKVKVFQGNQDSFTPVNSLDPPLTRYLRHPOSWVHQAIALRMEVLGCEAQDLY 1457
Db 1326 FQNGKVKVFQGNQDSFTPVNSLDPPLTRYLRHPOSWVHQAIALRMEVLGCEAQDLY 1383

RESULT 13
W11422
ID W11422 standard; Protein; 2342 AA.
AC W11422;
DT 20-NOV-1997 (first entry)
DE Active Factor VIII:C analogue, delta 1311-1320, + Pro insertion.

KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophiliac; therapy;
KW proteolytic cleavage.
OS Homo sapiens.
OS Synthetic.
FH Key location/Qualifiers
FT peptide 1..19
FT /note= "signal peptide"
FT protein 20..2342
FT /note= "mature Factor VIII:C"
FT region 20..1658
FT /note= "heavy chain fragment"
FT misc_difference 1329..1330
FT /note= "site of 10 residue deletion"
FT misc_difference 1330
FT /note= "inserted residue"
FT region 1659..2341
FT /note= "light chain fragment"
FT domain 760..1658
FT /note= "B domain"
PN W09703195-A1.
PD 30-JAN-1997.
PF 09-JUL-1996; U11444.
PR 11-JUL-1995; US-001025.
PA (CHIR) CHIRON CORP.
PI Cohen FE, Hung DT, Innis M;
DR WPI: 97-119050/11.
PT Factor VIII:C analog modified adjacent to a non-activating Arg
PT residue - used in the treatment of haemophiliacs, by improvement of
PT haemostasis
PS Claim 27; Page -: 90pp; English.
CC W11330-W11472 represent active Factor VIII:C analogues of the invention.
CC These sequences were created by mutating the wild type Factor VIII:C
CC coding sequence (see T51357) using mutagenic primers. The analogues
CC comprise a native Factor VIII:C polypeptide modified at a site adjacent
CC to a non-activating Arg residue so that a Arg-Pro or Pro-Arg dipeptide is
CC created. Factor VIII:C is a large glycoprotein that participates in the
CC blood coagulation cascade that ultimately converts soluble fibrinogen to
CC insoluble fibrin clot, effecting haemostasis. A deficiency in Factor
CC VIII:C is responsible for haemophilia A, which is an X-chromosome-linked
CC inherited bleeding diathesis. Factor VIII:C is activated by plasma
CC proteases, such as thrombin. During activation the mature polypeptide is
CC cleaved to generate heavy and light chain fragments that are further
CC cleaved. Complexes of two or more of the analogues, nucleic acids and
CC vectors encoding them may be used alone or in conjunction with each
CC other, for the prevention or treatment of active Factor VIII:C deficiency
CC in a mammal. The analogues may be used as immunogens to raise antibodies,
CC and in the treatment of haemophiliacs, by improvement of haemostasis. The
CC analogues are resistant to proteolytic cleavage and display increased
CC plasma half-life. They may be administered at lower dosages and by
CC different modes of administration.
SQ Sequence 2342 AA;

Query Match 94.2%; Score 7344.5; DB 1; Length 2342;
Best Local Similarity 62.2%; Pred. No. 0;
Matches 1457; Conservative 0; Mismatches 0; Indels 885; Gaps 1;
QY 1 MQIELSTCFLLCLRFCSATRRYYLGAVELSWDYMQSDLGELPYDARFPYPKSPFN 60
Db 1 MQIELSTCFLLCLRFCSATRRYYLGAVELSWDYMQSDLGELPYDARFPYPKSPFN 60
QY 61 TSVYKKTLFVERTDHLFNIAKRPWPWGLGPTIQAEVYDYVITLKNMASHPVSLHAV 120
Db 61 TSVYKKTLFVERTDHLFNIAKRPWPWGLGPTIQAEVYDYVITLKNMASHPVSLHAV 120
QY 121 GVSYWKASGAEYDDQTSQREKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTYSLSH 180
Db 121 GVSYWKASGAEYDDQTSQREKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTYSLSH 180
QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLAKFILLFAVPEGKSWHSETKNSLMQDRD 240
Db 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLAKFILLFAVPEGKSWHSETKNSLMQDRD 240

Db 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLLHKFILLFAVFDEGKSMHSETKNSLMODRD 240
QY 241 AASARAWPKMHTVANGYVNRSLPGLIGCHRSVYWHVIGMGTTPREVHSIFLEGHTFLVRNH 300
Db 241 AASARAWPKMHTVANGYVNRSLPGLIGCHRSVYWHVIGMGTTPREVHSIFLEGHTFLVRNH 300
QY 301 RQASLEISPTIFLTAQTLMDLGGFLLFCHISSHQHDMGMEAYVAVDSCPEEPOLBKKNE 360
Db 301 RQASLEISPTIFLTAQTLMDLGGFLLFCHISSHQHDMGMEAYVAVDSCPEEPOLBKKNE 360
QY 361 EAEDYDDDLTDEMDYVRFDDDNPSFIQIRSVAKKHPTWVHYIAAEEDWDYAPLVLA 420
Db 361 EAEDYDDDLTDEMDYVRFDDDNPSFIQIRSVAKKHPTWVHYIAAEEDWDYAPLVLA 420
QY 421 PDDRSYKSQYLNNGPQIRGRKYKVRMAYTDETFKTRAIQHSGLGPLLYGEVDTL 480
Db 421 PDDRSYKSQYLNNGPQIRGRKYKVRMAYTDETFKTRAIQHSGLGPLLYGEVDTL 480
QY 481 LIIFKNQASRPYMIYPHGITDVRLYSRRLPKGVKHLKDFPILPGEIEFKYKWTVVEDGP 540
Db 481 LIIFKNQASRPYMIYPHGITDVRLYSRRLPKGVKHLKDFPILPGEIEFKYKWTVVEDGP 540
QY 541 TKSDPRCLTRYSSFVNMERDLASGLIGPLLCYKESVDQRGNOIMSDKRNVLFSVEDE 600
Db 541 TKSDPRCLTRYSSFVNMERDLASGLIGPLLCYKESVDQRGNOIMSDKRNVLFSVEDE 600
QY 601 NRSWYLTENIQREFLPNPAGVQLEDEPEFOASNIMHSINGYVDSIQLSVCLHEVAYWYLS 660
Db 601 NRSWYLTENIQREFLPNPAGVQLEDEPEFOASNIMHSINGYVDSIQLSVCLHEVAYWYLS 660
QY 661 IGAQTDFLSVFESGYTFKHKNVYEDTLTLFPFSGETVFMSEMENPGLMILGCHNSDFRNRG 720
Db 661 IGAQTDFLSVFESGYTFKHKNVYEDTLTLFPFSGETVFMSEMENPGLMILGCHNSDFRNRG 720
QY 721 MTALLKVSSCDKMTGDYEDSYEDISAYLLSKNNAIEPRSF----- 762
Db 721 MTALLKVSSCDKMTGDYEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQENATTI 780
QY 762 ----- 762
Db 781 PENDIEKTDPFWAHRTMPKIQNVSSDDLMLLROSPTPHGLSLSDLQEAKEYTESDDPS 840
QY 762 ----- 762
Db 841 PGALDSNNSLSEMTHERPOLHSGDMVFTPESGIQLRLNEKLGTAATELKKLDFKVSST 900
QY 762 ----- 762
Db 901 SNNLISTIPSDNLAAGTNTSSLGPPSPMVHYDSQDLTLFGKKSPLTESGSPLSLEE 960
Db 901 SNNLISTIPSDNLAAGTNTSSLGPPSPMVHYDSQDLTLFGKKSPLTESGSPLSLEE 960
QY 762 ----- 762
Db 961 NNDKLLSEGLMNSQESSWGKNVSTESGRLFKGRAGHAPALLTKDNALFKVISILKTN 1020
QY 762 ----- 762
Db 1021 KTSNSATNRKTHIDPSLLIENSQVWQNLIESDTEFKKVTPLIHDRMLDKNATLRL 1080
QY 762 ----- 762
Db 1081 NHMSNKTSSKNMENVQOKKEGPIDPAQNPDMSFFKMLFLPESARWIOQTHGKNSLNSG 1140
QY 762 ----- 762
Db 1141 QGSPKQLVSLGPEKSVEGQNFLSEKNKVVGKEFTKDVGLKENVFPSSRNLFLTNLDN 1200
QY 762 ----- 762
Db 1201 LHENNTHQEKKIOEIEKKEKTLIQENVVLPQIHVITGTKNFMKNLFLSTRQNVESGYD 1260
QY 762 ----- 762
Db 1261 GATAPVLODFRSLNDSTNRTKKTHTAHFSKKEEENLEGLNQTKQIVEKYACTTRISPNT 1320

QY 762 ----- 762
Db 1321 SQQNFVTQRPLEELETELEKRIIVDDTSTOWSKNMKHLTPSTLLQIDYNEKEKAITOSPL 1380
QY 762 ----- 762
Db 1381 SDCLTRSHSIPQANRSPPLIAKVSFSPSIRPIYLTVLFQDNSSHLPAASYRKKDSGVQE 1440
QY 762 ----- 762
Db 1441 SSHFLOGAKKNNSLAILTLEMGTQDQREVSGLSGTSATNSVTYKKVENTVLPKPDLPKTS 1500
QY 762 ----- 762
Db 1501 KVELLPKVHIYQDLEPTETSNBPGHLDLVEGSLLOQTEGAIKWNANRPGKVPFLRYA 1560
QY 762 ----- 762
Db 1561 TESSAKTPSKLLDPLAMDNHYGQIIPKEWKSQEKSPKTAFFKKDTILSLNAGESNHAI 1620
QY 762 -----SONPVYLRKHOREITRTTLOSDQEEIDYDTISV 795
Db 1621 AAINEGONKPELEVYTWAKQGRTERLCSQNPVYLRKHOREITRTTLOSDQEEIDYDTISV 1680
QY 796 EMKKEDFDIYDEBENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHYLRNRAQSGSVPOF 855
Db 1681 EMKKEDFDIYDEBENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHYLRNRAQSGSVPOF 1740
QY 856 KKVYFOEFTDGSFTQPLYRGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSLSI 915
Db 1741 KKVYFOEFTDGSFTQPLYRGELNEHLGLGPYIRAEVEDNIMVTFRNQASRPYSFYSLSI 1800
QY 916 SYEEDQROGAEPKKNFVKPNETTYFWKVQHNAAPTKEFDCKAMAYFSDVDLEKDVHSG 975
Db 1801 SYEEDQROGAEPKKNFVKPNETTYFWKVQHNAAPTKEFDCKAMAYFSDVDLEKDVHSG 1860
QY 976 LIGPLLCHTNTLNPAGROVYQEFALFTTIDETKSWYFTENMERNCRAPCNIOMEDP 1035
Db 1861 LIGPLLCHTNTLNPAGROVYQEFALFTTIDETKSWYFTENMERNCRAPCNIOMEDP 1920
QY 1036 TFEKENYRFHAINGYIMDTLPGLVNAQDQIRIWMYLLSGSNENIHSIHFSGHVFTVRKKEE 1095
Db 1921 TFEKENYRFHAINGYIMDTLPGLVNAQDQIRIWMYLLSGSNENIHSIHFSGHVFTVRKKEE 1980
QY 1096 YKMALNYLPGEVETVEMLPKSAKIWRVECLIGEHLHAGMSTLFLVYSNKCQTPMGASG 1155
Db 1981 YKMALNYLPGEVETVEMLPKSAKIWRVECLIGEHLHAGMSTLFLVYSNKCQTPMGASG 2040
QY 1156 HIRDFQITASGOYQOWAPKLARLHYSGSINAMSTKEPFSWIKYVDLAPMIHGIKTQGAR 1215
Db 2041 HIRDFQITASGOYQOWAPKLARLHYSGSINAMSTKEPFSWIKYVDLAPMIHGIKTQGAR 2100
QY 1216 QKFSSLYISOFIIMYSLDGKKMQTYRGNSTGTLMVEFGNVDSGCIKHNIENPPIIARYIR 1275
Db 2101 QKFSSLYISOFIIMYSLDGKKMQTYRGNSTGTLMVEFGNVDSGCIKHNIENPPIIARYIR 2160
QY 1276 LHPTHYSIRSTLRBELMGCDLNSCSMPLGMEKSAISDAQITASSYFTNMFATWSPSKARL 1335
Db 2161 LHPTHYSIRSTLRBELMGCDLNSCSMPLGMEKSAISDAQITASSYFTNMFATWSPSKARL 2220
QY 1336 HLOGRSNAMPQVANNPKEMIAQVDFQKTMKVTGTQGVKSLLTSMYVKEFLISSQDGHQ 1395
Db 2221 HLOGRSNAMPQVANNPKEMIAQVDFQKTMKVTGTQGVKSLLTSMYVKEFLISSQDGHQ 2280
QY 1396 WTLFFQNGKVKVQGNQDSFTPVNLSLDPPLTPRYLRIPHQSWVHQIALRMEVJGCEAOD 1455
Db 2281 WTLFFQNGKVKVQGNQDSFTPVNLSLDPPLTPRYLRIPHQSWVHQIALRMEVJGCEAOD 2340
QY 1456 LY 1457
Db 2341 LY 2342

RESULT 14
W11432 standard; Protein: 2344 AA.
AC W11432;
DE 20-NOV-1997 (first entry)
KW Factor VIII:C analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
KW proteolytic cleavage.
OS Homo sapiens.
OS Synthetic.
FH Key
FT peptide 1..19
FT protein /note= "signal peptide"
FT /note= "mature Factor VIII:C"
FT region 20..2344
FT /note= "heavy chain fragment"
FT modified_site 1332..1333
FT /note= "site of 7 residue deletion"
FT region 1662..2343
FT /note= "light chain fragment"
FT domain 760..1661
FT /note= "B domain"
PN W09703195-AL.
PD 30-JAN-1997.
PE 09-JUL-1996; U11444.
PR 11-JUL-1995; US-001025.
PA (CHIR) CHIRON CORP.
PI Cohen FE, Hung DT, Innis M;
DR WPI; 97-119050/11.
PT Factor VIII:C analog modified adjacent to a non-activating Arg
PT residue - used in the treatment of haemophilias, by improvement of
PT haemostasis
PS Claim 29; Page -: 90pp; English.
CC W11330-W11472 represent active Factor VIII:C analogues of the invention.
CC These sequences were created by mutating the wild type Factor VIII:C
CC coding sequence (see T51357) using mutagenic primers. The analogues
CC comprise a native Factor VIII:C polypeptide modified at a site adjacent
CC to a non-activating Arg residue so that a Arg-Pro or Pro-Arg dipeptide is
CC created. Factor VIII:C is a large glycoprotein that participates in the
CC blood coagulation cascade that ultimately converts soluble fibrinogen to
CC insoluble fibrin clot, effecting haemostasis. A deficiency in Factor
CC VIII:C is responsible for haemophilia A, which is an X-chromosome-linked
CC inherited bleeding diathesis. Factor VIII:C is activated by plasma
CC proteases, such as thrombin. During activation the mature polypeptide is
CC cleaved to generate heavy and light chain fragments that are further
CC cleaved. Complexes of two or more of the analogues, nucleic acids and
CC vectors encoding them may be used alone or in conjunction with each
CC other, for the prevention or treatment of active Factor VIII:C deficiency
CC in a mammal. The analogues may be used as immunogens to raise antibodies,
CC and in the treatment of haemophilias, by improvement of haemostasis. The
CC analogues are resistant to proteolytic cleavage and display increased
CC plasma half-life. They may be administered at lower dosages and by
CC different modes of administration.
SQ Sequence 2344 AA;

Query Match 94.2%; Score 7343.5; DB 1; Length 2344;
Best Local Similarity 62.2%; Pred. No. 0;
Matches 1457; Conservative 0; Mismatches 0; Indels 887; Gaps 1;

QY 1 M0IELSTCFLLCLRFCSATRRYYLGAVELSMDYMSDGLGELPVDAFPFPRVPSPPFN 60
DB 1 M0IELSTCFLLCLRFCSATRRYYLGAVELSMDYMSDGLGELPVDAFPFPRVPSPPFN 60
QY 61 TSVVYKTLFVEFTDHLFNIAKPRPPNGLLPTIQAEVYDVTVTILKNMASHPSLHAY 120
DB 61 TSVVYKTLFVEFTDHLFNIAKPRPPNGLLPTIQAEVYDVTVTILKNMASHPSLHAY 120
QY 121 GVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHTYVQVLKENGPMASDPLCLTYSYLISH 180

DB 121 GVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHTYVQVLKENGPMASDPLCLTYSYLISH 180
QY 181 VDLVKDLNSGILGALLVCBREGSLAKEKQTOLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
DB 181 VDLVKDLNSGILGALLVCBREGSLAKEKQTOLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
QY 241 AASARAWPKMHTVNGVYVNSLPLGLIGCHKRSVYWHVIGMTTPREVHSTFLEGHTFLVRNH 300
DB 241 AASARAWPKMHTVNGVYVNSLPLGLIGCHKRSVYWHVIGMTTPREVHSTFLEGHTFLVRNH 300
QY 301 RQASLEISPIITFLTAQTLMDLGQFLFCHISSHQHDMGMAVYKVDSCPEEPQLRMKNNE 360
DB 301 RQASLEISPIITFLTAQTLMDLGQFLFCHISSHQHDMGMAVYKVDSCPEEPQLRMKNNE 360
QY 361 EAEDYDDDLTDEMDVVRFDNNSPSFIQIRSVAKKHPTWVHYIAAEEDMDYAPLVLA 420
DB 361 EAEDYDDDLTDEMDVVRFDNNSPSFIQIRSVAKKHPTWVHYIAAEEDMDYAPLVLA 420
QY 421 PDDRSYKSQYLNNGPQIRGRYKKVRFMAYTDETKTREAIQHESGILGPLLYGEVDTL 480
DB 421 PDDRSYKSQYLNNGPQIRGRYKKVRFMAYTDETKTREAIQHESGILGPLLYGEVDTL 480
QY 481 LIIFKNQASRPYNIYPHGITDVBPPLYSRRLPKGVKHLKDFPLLPGEIFKYKWTVTVEDGP 540
DB 481 LIIFKNQASRPYNIYPHGITDVBPPLYSRRLPKGVKHLKDFPLLPGEIFKYKWTVTVEDGP 540
QY 541 TKSDPRCLTRYSSFVNMERDIALSGLIGPLLCYKESVDQROQIMSDKRNVLFSVEDE 600
DB 541 TKSDPRCLTRYSSFVNMERDIALSGLIGPLLCYKESVDQROQIMSDKRNVLFSVEDE 600
QY 601 NRSWYLTENIQRFLEPNPAGVQLEDEPEFOASNMHSINGVYFDSIQLSVCLHEVAYWYILS 660
DB 601 NRSWYLTENIQRFLEPNPAGVQLEDEPEFOASNMHSINGVYFDSIQLSVCLHEVAYWYILS 660
QY 661 IGAQTDFLSVFFSGYTFKKHMYEDTLTFPFSGETVFMSENPGLMIIGCHNSDFERNRG 720
DB 661 IGAQTDFLSVFFSGYTFKKHMYEDTLTFPFSGETVFMSENPGLMIIGCHNSDFERNRG 720
QY 721 MTALLKVVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSF----- 762
DB 721 MTALLKVVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSF----- 762
QY 762 ----- 762
DB 762 ----- 762
QY 781 PENDIEKTDPMFAHRTPMPKIQNVSSDDLMLLRQSPTPHGLSLSDLQEAKEYTFSDPS 840
DB 781 PENDIEKTDPMFAHRTPMPKIQNVSSDDLMLLRQSPTPHGLSLSDLQEAKEYTFSDPS 840
QY 762 ----- 762
DB 762 ----- 762
QY 841 PGAIDSNNSLSEMTHTFRPQLHSGDMVFTPESGLQLRLNEKLTATATLKKLDFKVSST 900
DB 762 ----- 762
QY 901 SNNLISTIPSDNLAAGTDNTSSLGPPSPVHYDSQDLDTLLFGKKSPLTESGPIUSEE 960
DB 762 ----- 762
QY 961 NNDKLLSGLMNSQESSWGKNVSTESGRLLFKGRAHPALLTKDNALFKVISILKTN 1020
DB 762 ----- 762
QY 1021 KTSNNSATNRKTHIDPILLIENSPSVQNILSDTEFKKVTPLIHDRMLMDKNATLRL 1080
DB 762 ----- 762
QY 1081 NMSNKTSSKNMEMVQOKKEGPIPPDAQNPDMSFKMLFLPESARWIOPTHGKNSLNSG 1140
DB 762 ----- 762
QY 1141 QGSPKQVLVSLGPEKSVEGQNFLESKNKVYVGKEFTKDVGLKEMVFPSSRNLFLTINDN 1200
DB 762 ----- 762

Db 1201 LHENNTHNOEKKIQEEIEKKETLIQENNVLPQIHVTGTGKNFKNLFLSTRONVEGSYD 1260
QY 762 ----- 762
Db 1261 GAYAPVLQDFRSLNDSTNRKKTAAHFSKKGEEENLEGLNQTKQIVEKYACTTRISPNT 1320
QY 762 ----- 762
Db 1321 SQQNFVQSRKRPLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKEGAITQS 1380
QY 762 ----- 762
Db 1381 PLSDCLTRSHSIPQANRSPLPIAKVSSFPSIRPIYLTRVLFQDNSSHLPAASYRKKDSGV 1440
QY 762 ----- 762
Db 1441 QESSHFLQAKKNNLSLAILLEMTGDQREVSLGTSATNSVTYKKVENTVLPKPLPKT 1500
QY 762 ----- 762
Db 1501 SGKVELLPKVHIYOKDLFPETSNQSPGHLDIVEGSLQGTGAIKWNEANRPGKVPFLR 1560
QY 762 ----- 762
Db 1561 VATESAKTPSKLLDPLAWDNHYGTQIPKEEMWSQEKSPKTAFAKKKDTILSLNACESNH 1620
QY 762 -----SONPVLKRHOEITRTTLQSDQEEIDYDDTI 793
Db 1621 AIAAINEGONKREIEVTWAKQGTERLCSQNPVLRHOREITRTTLQSDQEEIDYDDTI 1680
QY 794 SVEMKKEDFDIYDEDEENQSPRSQKTRHYFIAAVERLMDYGMSSPHVLRNRAQSGSVP 853
Db 1681 SVEMKKEDFDIYDEDEENQSPRSQKTRHYFIAAVERLMDYGMSSPHVLRNRAQSGSVP 1740
QY 854 QFKKVVFOEFTDGSFTQPLYRGELNEHGLGPIYIRAEVEDNIMVTFRNQASRPYSFSS 913
Db 1741 QFKKVVFOEFTDGSFTQPLYRGELNEHGLGPIYIRAEVEDNIMVTFRNQASRPYSFSS 1800
QY 914 LISYEEDQROGAEPKRNFKVPNETKTYFWKVQHMAPTKDEFCKAWAYFSDVLEKDVH 973
Db 1801 LISYEEDQROGAEPKRNFKVPNETKTYFWKVQHMAPTKDEFCKAWAYFSDVLEKDVH 1860
QY 974 SGLIGPLVCHTNTLNPAGHROYTVQEFALFTTFDETQSWTFTEEMERNCRAPCNIQME 1033
Db 1861 SGLIGPLVCHTNTLNPAGHROYTVQEFALFTTFDETQSWTFTEEMERNCRAPCNIQME 1920
QY 1034 DPTFKENYRFHAINGYIMDTLPGLVMAQDQIRIKWYLLSMGSNENIHSHFSGHVFTVRKK 1093
Db 1921 DPTFKENYRFHAINGYIMDTLPGLVMAQDQIRIKWYLLSMGSNENIHSHFSGHVFTVRKK 1980
QY 1094 EEWKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGMA 1153
Db 1981 EEWKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGMA 2040
QY 1154 SGHIRDFOITASGOYQOWAPKRLARLHYSGSINAMSTKEPFSWIKVDLLAPMIHGIKTQG 1213
Db 2041 SGHIRDFOITASGOYQOWAPKRLARLHYSGSINAMSTKEPFSWIKVDLLAPMIHGIKTQG 2100
QY 1214 AROKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIARY 1273
Db 2101 AROKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIARY 2160
QY 1274 IRLHPTHYSIRSLRMLMGLCCLNSCSMPLGMSKAISDAQITASSSYFTNMFATWSPSKA 1333
Db 2161 IRLHPTHYSIRSLRMLMGLCCLNSCSMPLGMSKAISDAQITASSSYFTNMFATWSPSKA 2220
QY 1334 RLHLQGRSNAMRPOVNNPKFWLOYDFQKTMKVTVGTQGVKSLTSMYVKEFLISSQDG 1393
Db 2221 RLHLQGRSNAMRPOVNNPKFWLOYDFQKTMKVTVGTQGVKSLTSMYVKEFLISSQDG 2280
QY 1394 HQWTLFFONGKVKVYFQGNQDSFTPVVNSLDPPLLTRYLRIHPOSVWHQIALRMEVLGCEA 1453
Db 2281 HQWTLFFONGKVKVYFQGNQDSFTPVVNSLDPPLLTRYLRIHPOSVWHQIALRMEVLGCEA 2340

QY 1454 QDLX 1457
Db 2341 QDLX 2344

RESULT 15
W11410
ID W11410 standard; Protein; 2344 AA.
AC W11410;
DT 20-NOV-1997 (first entry)
DE Active Factor VIII:C analogue, delta 768-775, + Pro Insertion.
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic therapy;
KW proteolytic cleavage.
OS Homo sapiens.
OS Synthetic.
OS key
FH key
FT peptide 1..19 location/Qualifiers
FT protein /note= "signal peptide"
FT protein 20..2344
FT region /note= "mature Factor VIII:C"
FT region 20..1660
FT misc_difference 786..787 /note= "heavy chain fragment"
FT misc_difference 787 /note= "site of 8 residue deletion"
FT misc_difference 787 /note= "inserted residue"
FT region 1661..2343 /note= "light chain fragment"
FT domain 760..1660 /note= "B domain"
FN W09703195-A1.
PD 30-JAN-1997.
PF 09-JUL-1996; U11444.
PR 11-JUL-1995; US-001025.
PA (CHIR) CHIRON CORP.
PI Cohen FE, Hung DT, Innis M;
DR WPI: 97-119050/11.
PT Factor VIII:C analog modified adjacent to a non-activating Arg
PT residue - used in the treatment of haemophilias, by improvement of
PT haemostasis
PS Claim 25; Page -: 90pp; English.
CC W11330-W11472 represent active Factor VIII:C analogues of the invention.
CC These sequences were created by mutating the wild type Factor VIII:C
CC coding sequence (see T51357) using mutagenic primers. The analogues
CC comprise a native Factor VIII:C polypeptide modified at a site adjacent
CC to a non-activating Arg residue so that a Arg-Pro or Pro-Arg dipeptide is
CC created. Factor VIII:C is a large glycoprotein that participates in the
CC blood coagulation cascade that ultimately converts soluble fibrinogen to
CC insoluble fibrin clot, effecting haemostasis. A deficiency in Factor
CC VIII:C is responsible for haemophilia A, which is an X-chromosome-linked
CC inherited bleeding diathesis. Factor VIII:C is activated by plasma
CC proteases, such as thrombin. During activation the mature polypeptide is
CC cleaved to generate heavy and light chain fragments that are further
CC cleaved. Complexes of two or more of the analogues, nucleic acids and
CC vectors encoding them may be used alone or in conjunction with each
CC other, for the prevention or treatment of active Factor VIII:C deficiency
CC in a mammal. The analogues may be used as immunogens to raise antibodies,
CC and in the treatment of haemophilias, by improvement of haemostasis. The
CC analogues are resistant to proteolytic cleavage and display increased
CC plasma half-life. They may be administered at lower dosages and by
CC different modes of administration.
SQ Sequence 2344 AA;

Query Match 94.2%; Score 7343.5; DB 1; Length 2344;
Best Local Similarity 62.2%; Pred. No. 0;
Matches 1457; Conservative 0; Mismatches 0; Indels 887; Gaps 1;
QY 1 MOIELSTCFPLCLLRFCFSATRRYYLGAVELSDMYQSDLGELPVDARFPYKSPFN 60
|||||

QY	1334	RLHLOGRSNAWRPQVNNPKEWLQVDFOKTMTGVTTOGVKSLTSMYKEFLLISSQDG	1393
Db	2221	RLHLOGRSNAWRPQVNNPKEWLQVDFOKTMTGVTTOGVKSLTSMYKEFLLISSQDG	2280
QY	1394	HQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLTRYLRIRHPQSWVHQIALRMEVLGCEA	1453
Db	2281	HQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLTRYLRIRHPQSWVHQIALRMEVLGCEA	2340
QY	1454	QDLY 1457	
Db	2341	QDLY 2344	

Search completed: August 19, 1999, 15:59:09
Job time: 2446 sec

